Regulated Trafficking of a Composite Biopolymer in the Parasite *Giardia lamblia*: From Organelle Proteomics to Gene Knockouts

---

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Petra B. Wampfler
aus
Diemtigen BE

Promotionskomitee

Prof. Dr. Adrian B. Hehl
(Vorsitz und Leitung der Dissertation)

Prof. Dr. François Verrey
Prof. Dr. Thorsten Hornemann
Dr. Werner Kovacs

Zürich, 2014

Regulated Trafficking of a Composite Biopolymer in the Parasite *Giardia lamblia*: From Organelle Proteomics to Gene Knockouts

---

Faculty of Science University of Zurich

Life Science Graduate School – Microbiology and Immunology
PhD Program University of Zurich

PhD Thesis

Submitted by

Petra Wampfler

from Diemtigen BE, Switzerland

Thesis Advisors:

Prof. Dr. Adrian B. Hehl
Institute of Parasitology

Prof. Dr. François Verrey
Prof. Dr. Thorsten Hornemann
Dr. Werner Kovacs

Zürich, 2014

Table of Contents

PART I SUMMARY

1. Summary

As the leading cause for protozoal diarrhea worldwide, the small intestinal parasite *Giardia lamblia* (syn. *G. intestinalis*, *G. duodenalis*) is an important pathogen of humans and animals and bears responsibility for considerable morbidity and large economic losses for the livestock industry. The flagellated, motile trophozoites attach to epithelial cells in the small intestine of the host. Antigenic variation of surface coat proteins permits the parasite to outwit the host immune system, leading to persistent infection associated with watery diarrhea, abdominal pain, resorption disorders, and weight loss. Aside from its medical importance, *Giardia* has become a valuable model for both cell and evolutionary biologists. In course of a perfect adaptation to its ecological niche, this unicellular eukaryote has undergone massive reductive evolution which resulted in an even lower cellular complexity than that of the predicted last eukaryotic common ancestor (LECA). Consistently, most cellular processes in *Giardia* are simplified, and organelles are present in a highly reduced form, if not lost altogether. The minimized cellular environment in *Giardia* permits both the study of basic cell biological questions in minimal context as well as exploring eukaryotic diversity.

Parasite transmission of *Giardia* via the fecal-oral route requires the formation of an extracellular matrix, the cyst wall, conferring resistance to harsh environmental conditions. Considering its efficiency as a protective barrier, the structural composition of the unique *Giardia* cyst wall is surprisingly minimal consisting of but three paralogous cyst wall proteins (CWP1-3) and a β-(1,3)-N-Acetyl-galactosamine (GalNAc) homopolymer. To mediate secretion of the cyst wall material (CWM) in the absence of a steady-state Golgi, *Giardia* generates specialized organelles termed encystation-specific vesicles (ESVs) *de novo* at ER exit sites during encystation. As post-ER compartments dedicated to accumulation, modification, sorting, and regulated secretion of the newly synthesized CWM, ESVs feature striking parallels to the eukaryotic Golgi on a functional level. Being of critical importance for transmission to a new host and thus representing potential drug targets, ESVs and the CWM have been under investigation for many years. However, a detailed characterization of their structure, composition, neogenesis and functionality has been difficult due to the paucity of ESV-specific factors and the challenges posed by the highly reduced and diverged parasite genome.

In order to identify novel ESV-associated factors that may help characterize the nature of these specialized secretory organelles, we developed a novel approach combining flow cytometry-based organelle sorting (FAOS), shotgun proteomics and *in silico* data filtration to define the ESV proteome. Analysis and validation of the ESV organelle proteomic dataset provide strong evidence that CWP1-3 are the only highly abundant and stage-specifically expressed protein components of the cyst wall. Our data suggest that ESV neogenesis and maturation are primarily driven by the cargo proteins (CWP1-3) themselves with the assistance of only few, as yet unknown minor factors, if any, similar to the

formation of dense-core secretory granules (DCSG) in mammalian cells. Moreover, a strong physical connection between the ER and ESVs seems to be maintained until sorting and secretion of the CWM begins. Based on our findings and the known inherent properties of CWPs for self-aggregation, disulfide-based and isopeptide cross-linking, as well as carbohydrate (GalNAc) binding, we established three testable models for cargo-driven ESV neogenesis and maturation with CWP1-3 as the key driving-factors of the process.

The molecular underpinnings of these models require a detailed analysis of each CWP's function and contribution to the organelle neogenesis and maturation process. A straightforward approach to address this question comprise gene knock-out studies which are performed relatively easily in many other eukaryotic model organisms, but are currently not possible in *Giardia* due also to the parasite's tetraploid genome and the minimal repertoire of suitable drug resistance markers. Alternative strategies geared towards gene knock-down have been applied in *Giardia* but provide, at best, information on the biological effects of reduced cellular protein concentrations. However, this might be insufficient for the detailed functional analysis of highly abundant proteins such as CWP1-3 during encystation.

To overcome these limitations and to be able to manipulate the *G. lamblia* genome in a more effective manner we implemented the *E. coli* phage P1-derived Cre/loxP system in *Giardia*. The simple two-component system is based on the recombinase Cre (causes recombination) which recognizes specific, short DNA target sites termed loxP (locus of crossing over (x) in P1) and mediates recombination between them, resulting in excision, insertion, translocation or inversion of the interjacent DNA sequence. The system is already applied to manipulate the genome of other protozoa, and we were able to demonstrate the functional expression of the Cre recombinase in *G. lamblia* in this work. The application of the Cre/loxP system in *Giardia* allowed the removal of a previously inserted puromycin drug-resistance cassette from the *Giardia* genome and its recycling by using it in a subsequent, additional round of genetic manipulation in the same cell line.

Using this key concept for serial removal of all four alleles of a target gene in both nuclei, the Cre/loxP system paves the way towards the first *bona fide* gene-knock out in *G. lamblia*. We now propose a detailed model for the Cre/loxP-based generation of *cwp1-3* gene knock-outs. The elimination of these key components will hopefully provide insight into neogenesis and functionality of ESVs as well as secretory transport assembly of the protective surface composite biopolymer, and help to unravel the nature of these specialized and unique secretory organelles.

2. Zusammenfassung

Als Hauptverursacher von Protozoen-verursachter Durchfallerkrankungen weltweit ist der einzellige Darmparasit *Giardia lamblia* (syn. *G. intestinalis*, *G. duodenalis*) ein nicht zu unterschätzendes Pathogen von Mensch und Tier und trägt die Verantwortung für eine beträchtliche Krankheitsziffer und grossen Schaden in der landwirtschaftlichen Tierhaltung. Der begeisselte und motile Trophozoit heftet sich an das Dünndarmepithel seines Wirtes an, wobei eine kontinuierliche Variation der Oberflächenantigene es ihm erlaubt, das Immunsystem des Wirtes zu überlisten. Dies führt zu persistierenden Infektionen, welche mit wässrigem Durchfall, Unterleibsschmerzen, Resorptionsstörungen und Gewichtsverlust einhergehen. Neben seiner medizinischen Relevanz ist *Giardia* zu einem einzigartigen Modellparasiten für Forschung in der Zell- und Evolutionsbiologie geworden. Aufgrund der evolutionsbedingen, perfekten Anpassung an seine ökologische Nische mittels massiver zellulärer Reduktion weist *Giardia* eine noch einfachere zelluläre Komplexität auf als der letzte gemeinsame eukaryotische Vorgänger (LECA). Die Konsequenz dieser reduktiven Evolution manifestiert sich in einer vereinfachten Version der meisten zellulären Prozesse und Organellen, falls diese überhaupt vorhanden sind. Diese minimale zelluläre Ausstattung in *Giardia* ermöglicht das Studium von grundlegenden zellbiologischen Fragestellungen innerhalb eines stark vereinfachten zellulären Kontextes, aber auch die Erforschung eukaryotischer Diversität.

Um auf dem fäkal-oralen Weg auf einen neuen Wirt übertragen und den harschen Umweltbedingungen standhalten zu können, benötigt *Giardia* eine protektive extrazelluläre Matrix, die Zystenwand. Letztere ist angesichts ihrer effektiven Schutzfunktion überraschend einfach aufgebaut, besteht sie doch nur aus drei verwandten Zystenwand-Proteinen (cyst wall proteins 1-3, CWPs 1-3) und einem β-(1,3)-N-Acetyl-Galactosamin (GalNAc) Zucker. Da *Giardia* keinen permanenten Golgi-Apparat mehr besitzt, wird das Material für die Zystenwand (cyst wall material, CWM) während des Enzystierungsprozesses mittels spezialisierten Organellen, den Enzystierungs-Spezifischen Vesikeln (encystation-specific vesicles, ESVs) gereift und sekretiert. Letztere gehen aus spezialisierten Regionen des Endoplasmatischen Retikulums hervor (ER exit sites) und sind verantwortlich für die Akkumulierung, Modifikation, das Sortieren und die sequentielle Sekretion des neu synthetisierten CWM. In einem funktionellen Sinne erkennt man bei ESVs klare Parallelen zum klassischen eukaryotischen Golgi. Aufgrund ihrer Relevanz in der Parasitenübertragung von Wirt zu Wirt sind ESV Organellen potentielle Angriffspunkte für Therapeutika und werden deshalb seit vielen Jahren intensiv erforscht. Eine detaillierte Charakterisierung der Zusammensetzung, Entstehung und Selbstregulation der Organellen war bisher jedoch nicht möglich, da erst wenige ESV-spezifischen Faktoren bekannt sind und letztere aufgrund des stark reduzierten und divergierten Parasitengenoms nur schwer zu identifizieren sind.
Um unbekannte ESV-assoziierte Proteine zu identifizieren und die Eigenschaften dieser spezialisierten Sekretionsorganellen besser charakterisieren zu können, haben wir einen neuen experimentellen Ansatz

entwickelt. Eine Kombination aus Fluoreszenz-basiertem Organellen Sortieren (FAOS), massenspektrometrischer Proteinbestimmung (shotgun mass spectrometry), gefolgt von computergestützter Filtration der generierten Daten erlaubte es uns, die Proteinzusammensetzung von ESV Organellen zu definieren.
Die Analyse und Validierung der erhaltenen Daten weist stark darauf hin, dass CWP1-3 die einzigen hochabundanten und enzystierungs-spezifisch exprimierten Komponenten der Zystenwand sind. Unsere Daten deuten auf eine Entstehung und Reifung von ESVs primär durch die Eigenschaften der akkumulierten Proteine (CWPs) selber hin, die mit Hilfe von nur wenigen oder keinen zusätzlichen Faktoren gesteuert wird, ähnlich der Entstehung von Sekretions-Granula in Säugetierzellen. Zudem scheint eine direkte Verbindung zwischen dem ER und ESVs über die gesamte Zeit der Organellen-Reifung zu bestehen. Basierend auf unseren Resultaten und den bekannten Eigenschaften von CWP1-3 mit sich selbst zu aggregieren, kovalente Disulfid- und Isopeptidverbindungen zu bilden, sowie den Zystenwand-spezifischen GalNAc Zucker zu binden, haben wir drei testbare Modell-Szenarien für die Bildung und Reifung von ESV Organellen entwickelt, wobei CWP1-3 die Schlüsselfaktoren darstellen.

Ein experimenteller Test dieser Modelle beansprucht eine detaillierte Analyse der Funktion jedes CWPs, um dessen individuellen Beitrag zur Entstehung und Reifung der ESV Organellen zu bestimmen. Im Idealfall würde man diese Fragestellung mit der Entfernung der jeweiligen Gene (knock-out) vom *Giardia* Genom angehen. Obschon dieser Ansatz in vielen anderen eukaryotischen Modellorganismen zu realisieren ist, stellt er in *Giardia* aufgrund des tetraploiden Genoms in zwei Zellkernen und der wenigen geeigneten Selektionsmarkern ein vorläufig unüberwindbares Problem dar. Alternative Strategien mit dem Ziel der Unterdrückung der Proteinexpression (knock-down) wurden in *Giardia* zwar angewandt, resultieren aber bestenfalls in der Reduktion der zellulären Proteinkonzentration. Dies ist für eine funktionelle Analyse von stark exprimierten Proteinen, wie CWP1-3 während der Enzystierung, mit grosser Wahrscheinlichkeit nicht ausreichend.
Um dieses Hindernis zu überwinden und das Genom von *Giardia* effektiver manipulieren zu können, haben wir das Cre/loxP System des *E. coli* Phagen P1 in *Giardia* implementiert. Das einfache zwei-Komponenten System basiert auf der Cre (causes recombination) Rekombinase, welche bestimmte kurze DNA Sequenzen, die loxP (locus of crossing over (x) in P1) Sequenzen, erkennt und rekombiniert. Dies resultiert im Ausschneiden, Inserieren, in der Verlagerung oder im Umdrehen der dazwischenliegenden DNA Sequenz. Das Cre/loxP System findet bereits Anwendung in der Genommanipulation von anderen Protozoen, und wir haben nun zeigen können, dass die Cre-Rekombinase auch in *G. lamblia* funktionell exprimiert wird.
Die Anwendung des Cre/loxP Systems in *Giardia* ermöglichte uns, eine zuvor eingesetzte Puromycin-Resistenzkassette aus dem *Giardia* Genom zu entfernen, und diese in einer darauffolgenden weiteren Genommanipulation derselben Zelllinie nochmals als Selektionsmarker zu verwenden. Dieses System birgt den Schlüssel zur seriellen Entfernung ausgewählter Zielgene aus dem *Giardia* Genom und bringt

uns dem ersten Gen knock-out in diesem Parasiten einen entscheidenden Schritt näher. Basierend auf dem Cre/loxP System haben wir eine detaillierte Strategie für die Realisierung von *cwp1-3* Gen knock-outs in *G. lamblia* entwickelt. Die Eliminierung dieser Schlüsselkomponenten wird Einblick in die Entstehung und Funktionsweise von ESVs gewähren und helfen, die natürliche Beschaffenheit dieser spezialisierten und einzigartigen Sekretionsorganellen zu enthüllen.

PART II AIM OF THE THESIS

For transmission to a new host, *G. lamblia* depends on the synthesis and deposition of a protective extracellular composite biopolymer, the cyst wall. Three *Giardia* cyst wall proteins (CWPs 1-3) constitute the major protein components of this extracellular matrix and are, in the absence of a conventional Golgi apparatus in *Giardia*, secreted to the cell surface by stage-specifically induced organelles termed encystation-specific vesicles (ESVs). The correct modification and sorting of CWPs in ESVs and their regulated secretory trafficking are essential for the functional integrity of the protective wall and thus plays a critical factor in parasite transmission.

Although subject of investigation for many years, information on CWPs and their secretory organelles is sparse. A detailed systematic analysis has been prevented not only by the parasites tetraploid genome and the paucity of drug resistance markers which have disallowed the application of forward genetic approaches to study gene and protein function, but also by the parasites diverged and highly reduced genome that makes the identification of organelle-associated factors a considerable challenge.

The aim of my thesis was to investigate the protein composition of ESVs to detect novel factors associated with these organelles and thereby lay the groundwork to gaining insight into ESV neogenesis, functionality and stage-specific regulation. In an approach with a similar goal but using a different angle, I aimed to implement the *E. coli* phage P1-derived Cre/loxP system in *G. lamblia* to significantly enhance the spectrum of genome manipulations in *Giardia* and provide the key for future gene knock-out studies of CWP1-3, the ESV cargo proteins.

PART III INTRODUCTION

1. *Giardia lamblia*

1.1. *Giardiasis* and the *Giardia* life cycle

It was in his own watery stool, where the pioneering Dutch microscopist Antony Van Leeuwenhoek discovered *Giardia*, and he entered the history books in 1681 with the following description of *G. lamblia* trophozoites: "*I have at times seen very prettily moving animalcules ….. Their bodies were somewhat longer than broad, and their belly, which was flattened, provided with several feet, with which they made such a movement …"* (documented by [1]). Not until 26 years ago, experimental infection studies on human volunteers lead to the proof of human pathogenicity of this small intestinal parasite by fulfilling the Koch postulates [2]. In the meantime, *Giardia* was recognized as one of the leading causes for protozoal diarrhea worldwide, bearing responsibility for substantial morbidity and economic loss [3]. The flagellated parasite causes acute or chronic diarrhea associated with abdominal pain, nausea, malabsorption, and weight loss. The prevalence ranges from 2-7% in developed countries to 20-30% in most developing countries worldwide [4]; *Giardia*sis was included in the "WHO neglected disease initiative" in 2004 [5].

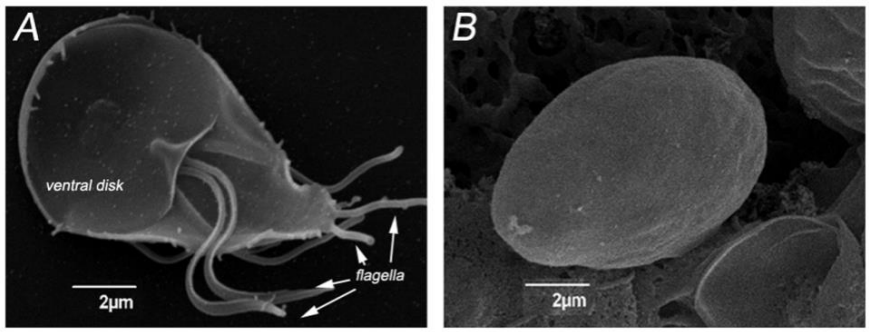


Figure 1: The two stages of the *G. lamblia* life cycle. **A)** *Giardia* trophozoite. The ventral disk structure and the flagella are indicated. **B)** *Giardia* cyst. Images: reference [6].

G. lamblia has a simple life cycle consisting of mobile, flagellated trophozoites (figure 1A) and environmentally resistant cysts (figure 1B) which are transmitted from host to host on the fecal-oral route. After ingestion by a new host and exposure to the acidic environment of the stomach, the *Giardia* cysts are triggered to initiate excystation [3]. The resulting trophozoites actively attach to the epithelium of the small intestine by a disk-like structure on their ventral side (figure 1A), and multiply asexually by binary fission. To evade the host immune system, trophozoites mediate antigenic variation in their protein surface coat [7]. Upon distinct environmental triggers, such as a change in bile concentration, bioavailability of lipids, or an increase in pH, trophozoites undergo a complex stage-differentiation process and transform to resistant cysts with low metabolic activity which are shed with the feces of the host. The *G. lamblia* life cycle, including encystation and excystation, can be reproduced axenically *in vitro* [8].

1.2. Evolutionary background

Until recently, *Giardia* was considered one of the earliest branching eukaryotes, together with other cytological simple parasitic lineages including *Trichomonas*, *Trypanosoma* and *Leishmania* (figure 2). Based on phylogenetic ribosomal RNA (rRNA) analysis and the apparent absence of several organelles, introns, and mitotic machinery, it was proposed that these lineages are primitively simple and branch at the very basis of the eukaryotic tree [9, 10].

During the last decade, more powerful algorithms and the increasing amount of available data for phylogenomic analysis has revolutionized the understanding of eukaryotic evolution. Previous phylogenetic trees, based on small subunit (16S-like) rRNA analysis which predicted eukaryotic evolution as a "progression" from the simple to the complex [9, 11], were recently replaced by an alternative model which predicts the existence of six major eukaryotic groups and removes the root of eukaryotes including the amitochondrial protists (figure 2) [10, 12-14]. The novel, egalitarian tree distributes protozoan parasites into three different eukaryotic major groups that are separated by several non-parasitic taxa. This entails that the parasitic lifestyle was "invented" independently on many separate occasions [11]. Hence, similar phenotypic variants of parasite life styles belonging to different taxa are the result of convergent evolution rather than inherited from an early eukaryotic ancestor. An example is provided by the antigenic variation of surface coat proteins, which is used by several protozoa as a strategy to evade the host immune system [15]. The phenomenon is well described in *Giardia* and *Trypanosoma* from the Excavata supergroup, as well as for *Plasmodium* belonging to the Chromalveolata [15]. Another example for convergent evolution is the synthesis of specialized organelles for the secretion of protective extracellular matrices [16], as observed in *Entamoeba invadens*, the social amoeba *Dictyostelium discoideum*, the ciliate *Tetrahymena termophila*, and *Giardia lamblia*, which belong to three different eukaryotic supergroups.

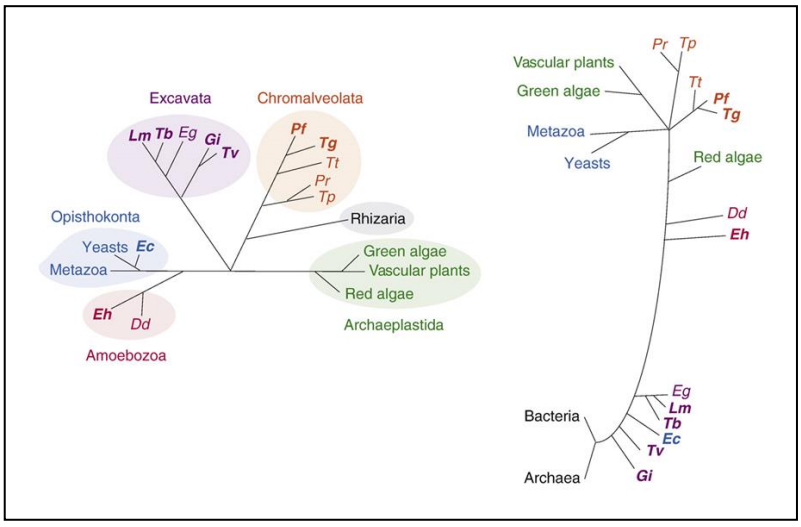


Figure 2: Different models for the evolution of eukaryotes [10]. **Left:** Recently proposed egalitarian (unrooted) tree based on molecular phylogenetic data [12-14]. The eukaryotes are categorized into six major groups. Parasitic protozoan lineages (bold) are dispersed in three groups. **Right:** Former elitist tree based on small subunit (16S-like) ribosomal RNA data only [9]. The amitochondrial protists represent the earliest diverging eukaryotic lineages. *Dd: Dictyostelium discoideum*; *Eh: Entamoeba histolytica*, *Gi: Giardia lamblia*; *Lm: Leishmania major*; *Pf: Plasmodium falciparum*; *Tg: Toxoplasma gondii*; *Tv: Trichomonas vaginalis*; *Tb: Trypanosoma brucei*.

Based on the current knowledge, the predicted last eukaryotic common ancestor (LECA) was not primitively simple but instead a "complete" eukaryotic cell which had a nucleus, an ER and Golgi apparatus, mitochondria, a complex eukaryotic cytoskeleton including flagella, and underwent mitosis and meiosis [12]. As a consequence, the apparent lack of mitochondria and other features, that originally characterized various eukaryotes as primitively simple, are now considered a result of minimization or secondary loss through reductive evolution. A case in point are the *Giardia* mitosomes [17], relic mitochondria that have lost many mitochondrial features such as the citric acid cycle, the respiration chain, and their genome (discussed below).

1.3. The *Giardia* organelle system

In course of perfect adaptation to its ecological niche, *Giardia* has undergone substantial reductive evolution. This resulted in a simplified version of most cellular processes, metabolic pathways, and minimization or even loss of cellular organelles such as mitochondria, peroxisomes, and a classical Golgi apparatus [11, 18, 19]. On a morphological and molecular level, three clearly identifiable organelle systems have been specified in *Giardia* trophozoites (figure 3) [20]: An extensive endoplasmic reticulum (ER), the peripheral vesicles (PVs), and relic mitochondrial organelles termed mitosomes. In encysting trophozoites, additional organelles termed encystation-specific vesicles (ESVs) are generated in a stage-specific manner, which contain secretory material destined for regulated export to the cyst wall [21, 22].

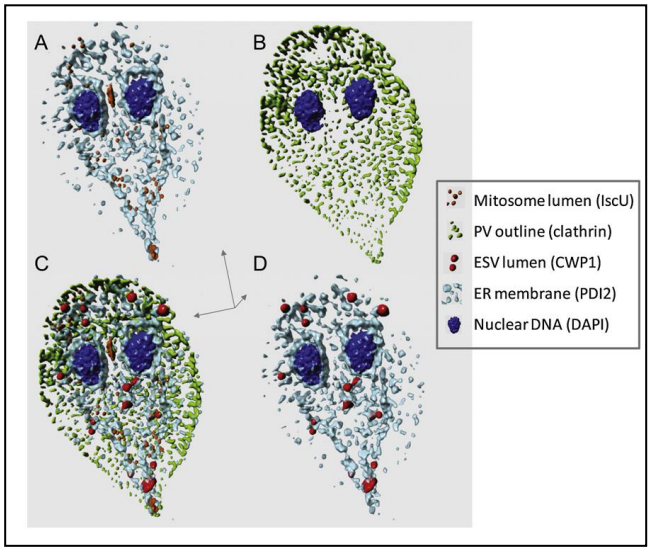


Figure 3: The three-dimensional model generated from confocal image stacks shows organelle systems in *G. lamblia* [20]. In *Giardia* trophozoites, three organelle systems have been identified: the mitosomes (orange, A, C), the peripheral vesicles (PVs, green, B, C), and the ER including the perinuclear region and nuclear envelopes (light blue, A, C, D). In encysting cells, an additional set of organelles termed encystation-specific vesicles (ESVs) is established *de novo* (red, B, C). Nuclear DNA is stained with DAPI. Organelle-specific markers are *IscU:* Iron-sulfur-cluster assembly protein, *CWP1:* cyst wall protein 1, *PDI2:* protein disulfide isomerase 2.

The endoplasmic reticulum (ER)

The *Giardia* ER consists of cisternal and tubular elements and extends from the nuclear envelopes throughout the entire cell body [23] and touches but does not penetrate the cortical space occupied by PVs [20]. With respect to secretory protein trafficking, *Giardia* possesses a conventional ER: conserved

machinery for translocation mediates co-translational import of secreted proteins into the ER, and protein folding is assisted by a series of conserved chaperones such as protein disulfide isomerases (PDIs), the Hsp70 Binding Protein BiP, and peptidyl-prolyl *cis-trans* isomerases [11]. Unlike in many eukaryotes, N-glycosylation of proteins is restricted to the addition of $GlcNAc_{1-2}$ to asparagine [24]. Consistent with this, the parasite lacks the ER based (calnexin/calreticulin) quality control system for glycoprotein folding, and encodes neither enzymes required for synthesis of the conventional eukaryotic core oligosaccharide $GlcNAc_2Man_9Glc_3$ nor factors for N-glycan or O-glycan processing in the ER or Golgi [11].

The ER is positioned at the crossroads of endocytic and exocytic protein trafficking in *Giardia*. Due to the absence of a conventional Golgi apparatus, the ER plays a major role in direct secretion of proteins to their target organelles or the plasma membrane [21]. In addition, the organelle is involved in endocytic trafficking by uptake and catabolism of pre-sorted bulk endocytosed material from the *Giardia* peripheral vesicles [25].

Peripheral vesicles (PVs)

Peripheral vesicles (PVs) are small, 150 nm-sized organelles that underlie the plasma membrane of the entire dorsal side and a specialized region termed bare zone at the center of the ventral disk [26]. PVs mediate endocytosis of soluble and membrane-bound factors and acidify [25-28]. The organelles periodically open to the environment, presumably by direct fusion with the plasma membrane, and take up fluid phase material non-selectively [25, 26, 29]. While most fluid phase markers remain in PVs until extruded upon the next cycle of fusion with the plasma membrane, selected markers are transported rapidly inward to the proximal ER and the nuclear envelope [25], suggesting a discriminatory sorting function associated with PVs. Besides their endocytic and sorting characteristics [30-32], PVs have a lysosomal nature since they acidify and become digestive compartments [29]. Containing cathepsins and acid hydrolases, PVs are likely involved in degradation of bulk endocytosed material before their trafficking to the ER [25]. Further, PVs might play a role in secretory protein trafficking as indicated by various studies. The organelles were suggested to secrete distinct proteins including an acid phosphatase required for decomposition of the cyst wall during excystation [33], and also variant-specific surface proteins (VSPs) which have been detected in PVs in immuno-electron microscopy studies [32, 34].

Mitosomes

The *Giardia* mitosomes are less than 200 nm-sized, double-membrane bounded cytoplasmic compartments. The organelles come in two flavors: they are either distributed throughout the cell (peripheral mitosomes) or tightly clustered in the basal body region between the two nuclei (central mitosomes)[35]. While the *Giardia* mitosomes harbor a simple protein import and folding machinery [35-38] and function in synthesis and assembly of iron-sulfur clusters [17, 38-40], they lack other classical

mitochondrial attributes such as cristae, the respiration chain, the citric acid cycle, or a mitochondrial genome.

Encystation specific vesicles (ESVs)

The process of encystation in *Giardia* includes synthesis and regulated secretion of a polymerizing extracellular matrix, the cyst wall (CW). Considering its high environmental resistance, the *Giardial* CW is a biopolymer of surprisingly low complexity. Three paralogous cyst wall proteins (CWP1-3) are the only known structural proteins of the extracellular wall [41-43]. The CWPs are between 241 and 362 amino acids in length and share high sequence identity. They all contain predicted hydrophobic N-terminal secretory signal peptides, a central region consisting of 4 to 5 tandem arrayed leucine-rich repeats, followed by cysteine-rich regions (figure 4). CWP2 contains an additional C-terminal domain of 121 residues which is rich in basic amino acids and is not found in CWP1 or CWP3. This C-terminal part of this basic extension is proteolytically processed during ESV maturation [44] (discussed below). In addition to the three CWPs, the cyst wall contains a β-(1-3)-GalNAc polymer which is unique to *Giardia* and makes up about 60% of the wall material [45, 46]. Where the sugar polymer is synthesized and how and when it is trafficked to the cell surface has not been resolved definitively.

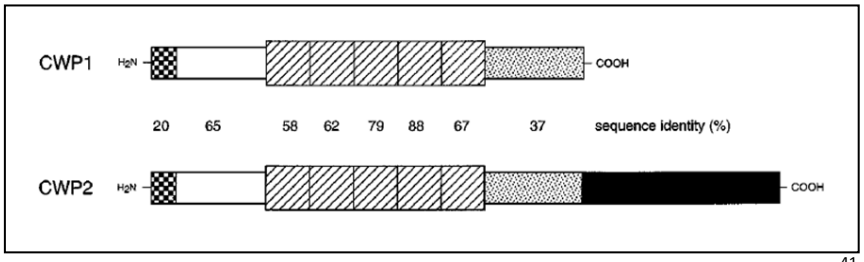


Figure 4: Schematic depiction of CWP1 and CWP2 domain organization [41]. Checkered boxes: signal peptides, cross-hatched boxes: leucine rich repeats, stippled boxes: cysteine-rich regions, black box: 121 residue C-terminal domain. CWP3 is very similar to CWP1 and is not shown.

A very unusual aspect is that *Giardia* has no steady-state Golgi apparatus with which to organize secretory transport of proteins and regulated secretion of the cyst wall material (CWM) in particular. Instead, encystation-specific vesicles (ESV) are formed *de novo* from ER exit sites during encystation [21, 47, 48]. As post-ER compartments, these organelles accumulate and delay only the newly synthesized CWPs and no other secretory cargo for post-translational modification. In addition, ESVs play a key role in sorting and sequential secretion of the *Giardia* CWM to the cell surface where it polymerizes [44]. Hence, ESV organelles closely resemble the eukaryotic Golgi on a functional level. In fact, the latest data strongly support the hypothesis that ESVs are inducible, transient Golgi cisterna-equivalents: COPII-dependent ESV neogenesis appears to be highly analogous to formation of *cis*-Golgi cisternae [47, 48], and maturing ESVs sort cargo in a manner that is very reminiscent of sorting at the *trans*-Golgi network [44]. Further,

ESVs are sensitive to Brefeldin A, a fungal metabolite that causes Golgi-disassembly in higher eukaryotes [22, 49]. ESVs were proposed to be transient (i.e. pulsed) single-stacked Golgi-analogs that undergo maturation from a *cis*- to a *trans*-stage, reminiscent of the cisternal maturation model for Golgi function in higher eukaryotes [21, 44]. On the other hand, ESVs are not a permanent feature of trophozoites and clearly differ from a classical Golgi from a structural point of view, since neither conserved structural nor functional proteins of the Golgi have been identified in ESVs so far.

1.4. Constitutive and regulated protein secretion in *Giardia lamblia*

As a consequence of secondary reduction, *Giardia* trophozoites have no classical Golgi apparatus, and secreted proteins are trafficked from the ER straight to the target organelle or the plasma membrane [21, 50, 51]. A large part of the constitutive secretory activity of trophozoites includes the secretion of membrane-spanning variant-specific surface proteins (VSPs) [20]. These proteins with cysteine-rich exodomains mediate antigenic variation on the trophozoite plasma membrane, helping the parasite to evade the host immune system [7]. Of the predicted 235 to 275 VSPs [11], only one coats the parasite membrane at a time [52]. Antigenic switching, i.e. exchange of the VSP type currently expressed on the surface by a different one, occurs every 6 to 13 generations [53] and is dependent on RNA interference (RNAi) mediated silencing mechanisms as discussed below [54]. The VSPs extracellular domains are cleaved and exodmaons are released as soluble antigens into the environment [55]. This results in continuous secretion and turnover of the currently expressed VSP type on the plasma membrane. Surface targeting of trafficked VSPs from the ER requires an absolutely conserved C-terminal CRGKA motif [21]. Trafficking likely occurs directly from the ER to the plasma membrane [21, 50, 51], and is sensitive to Brefeldin A [41, 49], a fungal metabolite interfering with the functions of the GTPase Arf1. This suggests involvement of COPI-dependent membrane carriers in VSP export (figure 5).
In addition to VSPs, a group of non-variable cysteine-rich proteins [56] are also targeted to the plasma membrane, as well as an unknown number of the roughly 500 proteins that are predicted to enter the secretory pathway [20] (http://*Giardia*db.org/*Giardia*db/showQuestion.do?questionFullName=GeneQuestions.GenesWithSignalPeptide).

In cells which differentiate to cysts, a regulated secretory pathway is generated by *de novo* formation of specialized organelles termed encystation-specific vesicles (ESVs)[21, 22]. ESVs secrete the CWM to the parasite surface (figure 5). Whether the GalNAc-sugar component of the cyst wall is trafficked via ESVs to the cell surface or takes an alternative route is currently unknown.

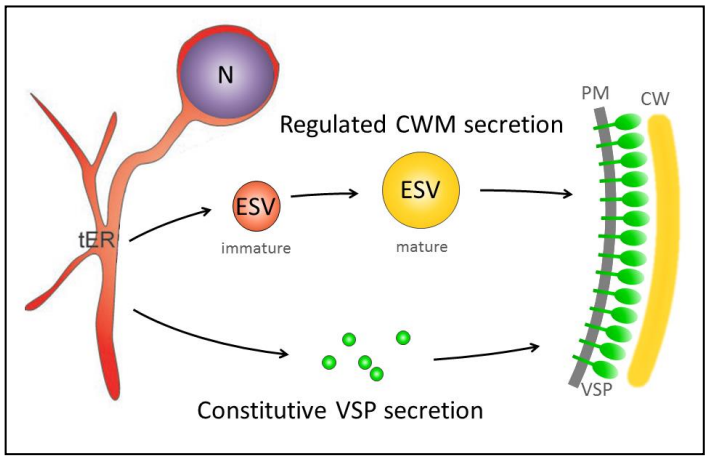


Figure 5: Basic model of secretory pathways in *G. lamblia.* **Top:** Regulated secretion of CWM in encysting trophozoites. The CWM is exported from transitional ER (tER) sites in a COPII-dependent manner and accumulates in early ESV organelles (red) which are established *de novo*. ESVs grow in size by accumulation of CWM and undergo a maturation process. From mature ESVs (yellow), the CWM is secreted via unknown transport intermediates to the cell surface, where it polymerizes to from the cyst wall. **Bottom:** Constitutive secretion of variant-specific surface proteins (VSPs) in trophozoites and encysting cells. VSPs are directly trafficked from tER sites to the plasma membrane, likely in COPI-dependent transport intermediates. *tER:* transitional ER; *N:* nucleus; *ESVs:* encystation-specific vesicles; *PM:* plasma membrane; *CW:* cyst wall; *VSPs:* variant specific surface proteins.

A frequently used laboratory protocol to induce stage-differentiation of *G. lamblia* trophozoites is a two-step method based on bile deprivation and a subsequent increase in pH combined with addition of porcine bile [57]. Expression of CWPs is induced stage-specifically from a zero level in trophozoites, and mRNA levels peak at 7 hours post induction (p.i.) [44, 57]. Additional modulation of CWP expression is dependent on a conserved motif in their promoter sequence which serves as binding region for a Myb1-like transcription factor [57]. In addition to CWP1-3, *Giardia* Myb1 positively regulates transcription of a set of key enzymes required for the cyst wall sugar synthesis [57, 58].

Synthesis and co-translational import of CWPs into the ER lumen occurs during the first 7 hours of encystation (figure 6). The CWPs are selectively sorted out from constitutively secreted material at ER exit sites and leave the ER in a COPII dependent manner [21, 47, 48, 59]. Accumulation of CWPs in emerging ESVs is observed by fluorescence microscopy from 2 hours post induction (p.i.) of encystation and is completed after 8 to 10 hours p.i. [49]. The material is delayed in ESVs for several hours, allowing post-translational maturation; the highly homologous CWPs are partitioned into two fractions which are then secreted sequentially [44]. In a first step (8 to 12 hours p.i.), CWP3 and the proteolytically cleaved C-terminal fragment of CWP2 (ΔC) undergo selective condensation, resulting in the formation of two biophysically distinct CWM fractions inside ESVs. The remaining two components, CWP1 and the N-terminal part of processed CWP2 (ΔN), remain fluid and circulate between ESVs within a dynamic, tubular membrane possibly via the ER [47]. In a second phase (16 to 20 hours p.i.), the fluid components are sorted away from the condensed material and appear in small compartments near the cell periphery, before they are secreted rapidly and form a structurally resistant matrix on the cell surface. The condensed material remains in internal compartments, is decondensed and secreted slowly during several hours, and most likely forms an inner layer of the cyst wall [44].

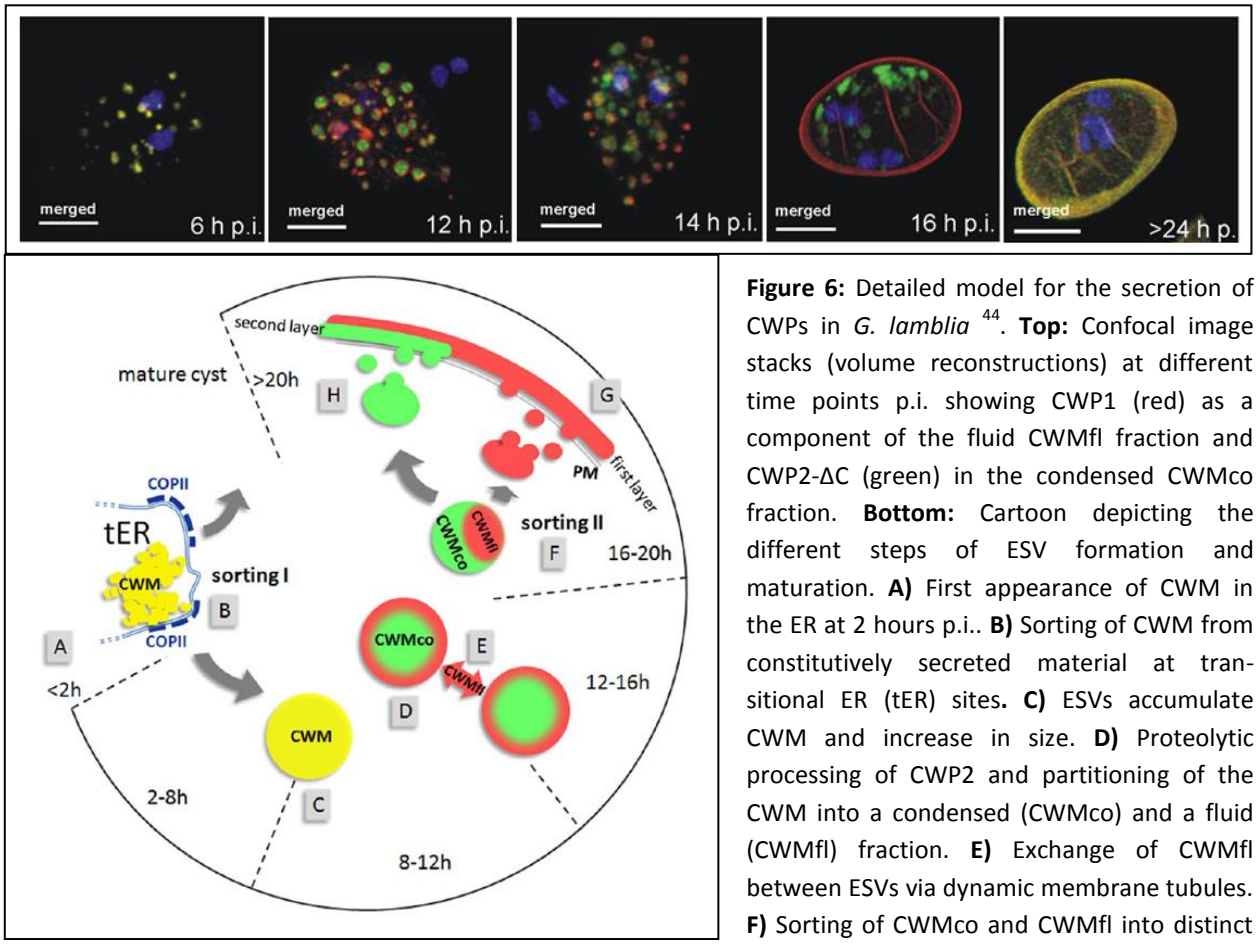


Figure 6: Detailed model for the secretion of CWPs in *G. lamblia* [44]. **Top:** Confocal image stacks (volume reconstructions) at different time points p.i. showing CWP1 (red) as a component of the fluid CWMfl fraction and CWP2-ΔC (green) in the condensed CWMco fraction. **Bottom:** Cartoon depicting the different steps of ESV formation and maturation. **A)** First appearance of CWM in the ER at 2 hours p.i.. **B)** Sorting of CWM from constitutively secreted material at transitional ER (tER) sites**. C)** ESVs accumulate CWM and increase in size. **D)** Proteolytic processing of CWP2 and partitioning of the CWM into a condensed (CWMco) and a fluid (CWMfl) fraction. **E)** Exchange of CWMfl between ESVs via dynamic membrane tubules. **F)** Sorting of CWMco and CWMfl into distinct compartments. **G)** Rapid secretion of CWMfl and establishment of an outer wall. **H)** Slow secretion of CWMco, likely involving decondensation. Maturation must be completed before cysts become water-resistant. *CWM:* cyst wall material; *co:* condensed; *fl:* fluid; *p.i.:* post induction of encystation; *tER:* transitional ER; *h:* hours post induction of encystation.

Besides CWP1-3, constituting the only known stage-specifically expressed CW components, a number of constitutively expressed minor proteins appear to be trafficked via ESVs to the cell surface. They include the high cysteine non-variant cyst protein (HCNCp) [56] and six EGF-like cyst proteins (EGFCP1-6) [60]. Whilst these proteins localized to the nuclear envelope and the peripheral ER in trophozoites, they were detected also in ESVs during encystation, and in the cell body and cyst wall of cysts [56, 60]. A calcium-binding protein termed granule specific protein (GSP) was detected in ESVs during encystation but not in the cyst wall [61]. GSP was suggested to regulate calcium-dependent degranulation of ESVs during encystation [61].

In total fewer than 20 proteins are known to re-localize stage-specifically from other cellular compartments and undergo a transient association with ESVs. Among these are several trafficking proteins either known or supposed to be involved in CWP transport from ER to ESVs during early

encystation (*Giardia*l homologs of Sar1, Sec31, Yip1, Rab1a), or in secretion of the CWM to the cell surface at later encystation time points (dynamin-like protein DLP, Arf1, Rab11) [18, 29, 47, 62, 63]. Proteasome complexes are recruited to ESV membranes during early encystation, perhaps as a result of post-ER quality control and associated degradation processes [62]. In addition, recruitment of COPI-coat components to ESVs and the KDEL-dependent retrieval of the ER chaperone Hsp70/BiP from ESVs suggested the presence of a COPI-based retrograde trafficking from ESVs to the ER [62].
Several SNARE (soluble N-ethylmalemide–sensitive factor attachment protein receptor) proteins were detected at ESV organelles [63]. Finally, two cysteine proteases are recruited to ESVs; a subtilisin-like proprotein convertase (SPC), localizing to the perinuclear ER in trophozoites [64], and a cathepsin-B like cysteine protease 2 (CP2) being one of several candidates discussed regarding the proteolytic processing of CWP2 [65].

While the CWPs reach the surface by trafficking through ESVs, it is still unclear when and where in the exocytic pathway the cyst wall sugar (CWS) component, a β-(1,3)-GalNAc homopolymer, is synthesized and how it is finally incorporated into the cyst wall [45, 46]. The recent literature suggests synthesis of the cyst wall monomer, UDP-GalNAc, from glucose by a series of cytosolic enzymes that are upregulated after induction of encystation [58]. The UDP-GalNAc monomer is polymerized into the β-(1-3)-GalNAc homopolymer by a putative cyst wall synthase [66]. Whereas the activity of the enzyme has been demonstrated [66], its identity was never unraveled.

1.5. Genetic organization in *Giardia lamblia*

The Giardia genome

The most widely used laboratory strain of *G. lamblia* (WB clone 6, ATCC 50803, human isolate [67]) has a compact genome of 11.7 MB in size which is fully sequenced [11, 68]. It codes for 6470 predicted open reading frames (ORFs) that frequently overlap and are distributed on 5 chromosomes [11]. Although the *Giardia* genome is size-wise comparable to that of yeast, *Giardia* has significantly simplified variants of most cellular processes and machineries, including those for transcription and translation, compared to those of yeast and other eukaryotic organisms [11]. The presence of many genes with bacterial or archaeal origins in *Giardia* suggests that lateral gene transfer has strongly influenced the shaping of the parasite's genome [11].

Transcription, mRNA processing, and translation

The promoter sequences in *Giardia* are unusually short with a length of 70 bp or less[69]. Although the *Giardia* transcription factors sometimes remarkably diverge from those of other eukaryotes [70], mRNA maturation in *Giardia* involves many features typical for eukaryotes including 5' capping [71], polyadenylation [11, 69], and rarely also intron splicing [11].

Transcripts in *Giardia* have short untranslated regions (UTRs), ranging from 0 to 14 nucleotides for 5'UTRs, and 10 to 30 nucleotides for 3'UTRs [69]. Based on the minimal 5'UTRs and some missing translation initiation factors it was proposed that initiation of translation in *Giardia* does not include ribosome scanning as observed in eukaryotic cells, but instead proceeds similar to prokaryotes [11]. It was found that each of the *Giardia* ORFs contains a so-called downstream box (DB) consisting of 13 nucleotides which is also found in *E. coli*. Due to a complementary region in the 16S ribosomal RNA, this conserved sequence promotes efficient initiation of protein synthesis even in the absence of a Shine Dalgarno sequence [11].

Only 5 genes containing introns were detected so far, and *cis*-splicing has been experimentally demonstrated for 4 of them [11, 72-74]. *Giardia* encodes a complete eukaryotic splicing machinery [11]. Since the number of introns in the parasite is very low, the question comes up why mRNA splicing is actually maintained in *Giardia*. A recent study reported the *trans*-splicing of two parts of the cytosolic chaperone Hsp90 mRNA in *Giardia* which is encoded by separate genes [75-77]. Involvement of the spliceosome as a key factor in the expression of essential housekeeping genes might explain the maintenance of the spliceosome.

RNA interference in Giardia

The eukaryotic RNA interference pathway plays a key role in gene silencing and consists of the three key factors Dicer, Argonaute, and an RNA-dependent RNA polymerase (RdRP). In contrast to higher eukaryotes, *Giardia* encodes only for a single copy of each factor [78-80]. The *Giardia* genes for Dicer and Argonaute were initially not identified due to the high divergence or even absence of normally conserved protein domains. Proof for the existence of a functional RNAi machinery in *Giardia* was provided by a recent study demonstrating that antigenic switching of *Giardia* VSP surface proteins is controlled by RNAi-based silencing [54]. Besides involvement in antigenic variation, the RNAi pathway seems to play an important role in the parasite since down-regulation of *Giardial* Argonaute leads to inhibition of cell growth [81].

The Giardia nuclei

As a member of the Diplomonads, *Giardia* has two diploid nuclei which make the cell effectively tetraploid [82, 83]. Both nuclei are transcriptionally active, and nuclear inheritance to daughter cells during semi-open mitosis occurs via equatorial segregation (figure 7)[82-84].

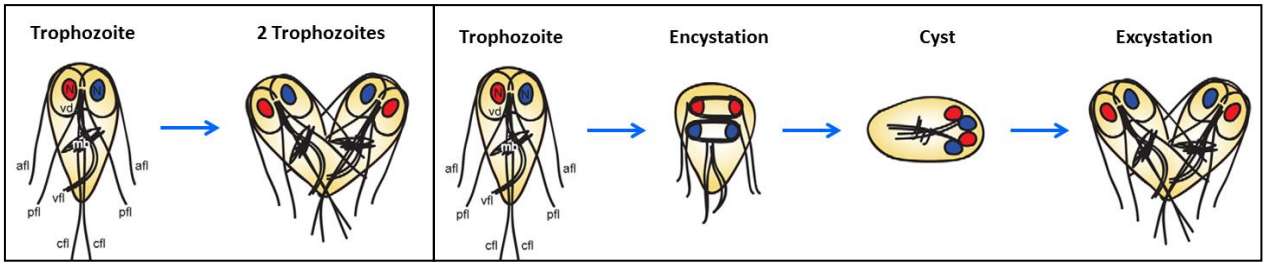


Figure 7: Nuclear inheritance in the *Giardia* life cycle [84]. **Left:** Trophozoite dividing by binary fission. The nuclei are inherited equatorially, so that each daughter cell receives one copy of each parental nucleus. **Right:** Trophozoite undergoing encystation. An incomplete mitosis results in a cyst with four daughter nuclei containing four genome complements each. During excystation, the emerged exyzoite undergoes two rounds of cytokinesis and one round of karyokinesis, resulting in four daughter trophozoites. Each of them receives one copy of the initial parental nuclei.

Until recently, *Giardia* has been considered a strictly asexual organism replicating only by binary fission. However, the unexpectedly low level of heterozygosity raised the question how homozygosity between the separated nuclei as well as on the population level is maintained [11]. Possible answers were found only recently through the detection of meiotic genes and evidence from population genetic studies which support the occurrence of sexual recombination in *Giardia* [85-87]. In addition, it was observed that *Giardia* nuclei can exchange chromosomal material in cysts (figure 7) [84]. These findings might explain both the genome homogenization on the population level and between to two nuclei.

Giardia's binuclear state raises interesting questions about the emergence and functionality of the two nuclei, as well as how orchestration of transcription is controlled. On the other hand, the tetraploid genome organization strongly limits the application of genetic and cell biological approaches to investigate this intestinal parasite. The situation is worsened by the limited repertoire of only two drug-resistance markers suitable for selection of transgenic *Giardia* [88-91]. Accordingly, genetic manipulations of relatively low complexity, such as gene knock outs, are easily realized in many other eukaryotic organisms but constitute a currently insuperable barrier in *Giardia*.

2. The Cre/loxP system

In site-specific recombination (SSR), a DNA molecule is cut at two specific positions by a recombinase, and the cut ends are recombined differently [92]. SSR systems are found in many organisms where they play diverse roles such as phage integration into and excision from host cell genomes, plasmid copy number regulation, or monomerization of multimeric bacterial plasmids and chromosomes [92]. The best investigated SSR systems are the ones from bacteriophages, bacteria, and yeast. The DNA recombination sites of different systems vary considerably in size, ranging from 30 bp up to 240 bp. Many SSR systems require the help of accessory proteins in addition to the recombinase [92].

The *E. coli* P1 phage-derived SSR system is a simple bipartite system and based on short recombination sites. The Cre-recombinase (causes recombination) mediates inter- and intramolecular recombination of short 34 bp-sized loxP sites (locus of crossing over (x) in P1) without the involvement of any co-factors [93, 94]. Depending on the relative orientation of two loxP sites with respect to one another, Cre-mediated recombination results in excision, insertion, inversion, or translocation of loxP-flanked ("floxed") DNA sequences (figure 8) [95].

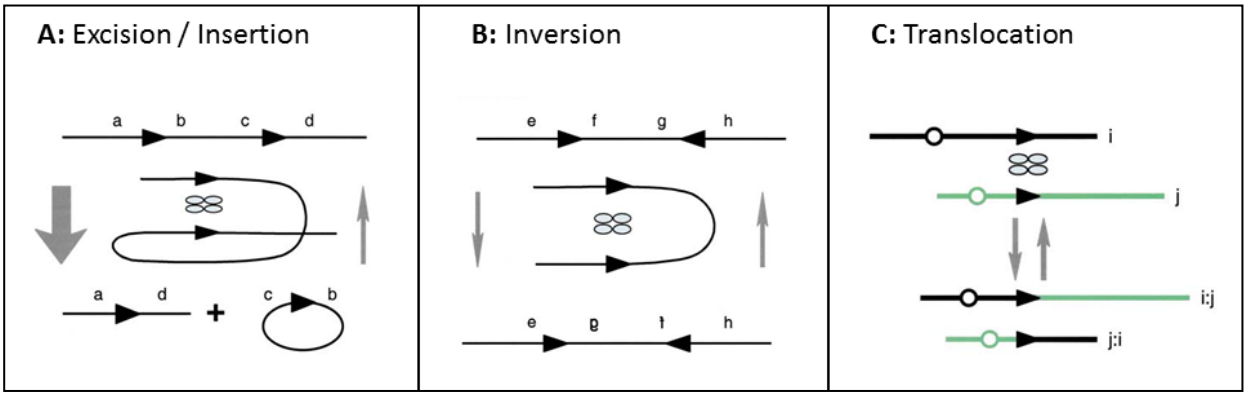


Figure 8: Recombination reactions mediated by Cre [95]. The outcome depends on the relative orientation of the loxP sites with respect to one another. **A)** Identical orientation of loxP sites: Cre excises a floxed sequence which results in a circular molecule (left) or integrates a circular molecule into a linear molecule each possessing a loxP site (right). **B)** Inverted orientation of loxP sites: Cre mediates inversion of floxed DNA sequence. **C)** LoxP sites present on different linear molecules: Cre exchanges sequences distal to loxP sites. This intermolecular recombination can occur between loxP sites on non-homologous chromosomes. *Arrowheads:* loxP sites; circle: Cre momoner; *a-j:* distinct DNA sequences.

The 34 bp loxP site consists of an asymmetrical core which is flanked by two inverted repeats (figure 9). To catalyze the recombination reaction, 4 Cre recombinase monomers bind to the inverted repeats of two loxP sites (figure 9). Interestingly, the Cre-recombinase tolerates certain variations in the targeting site. The alternative lox sites are still recognized by Cre, but recombination occurs only between homotopic pairs of these variant sites [95].

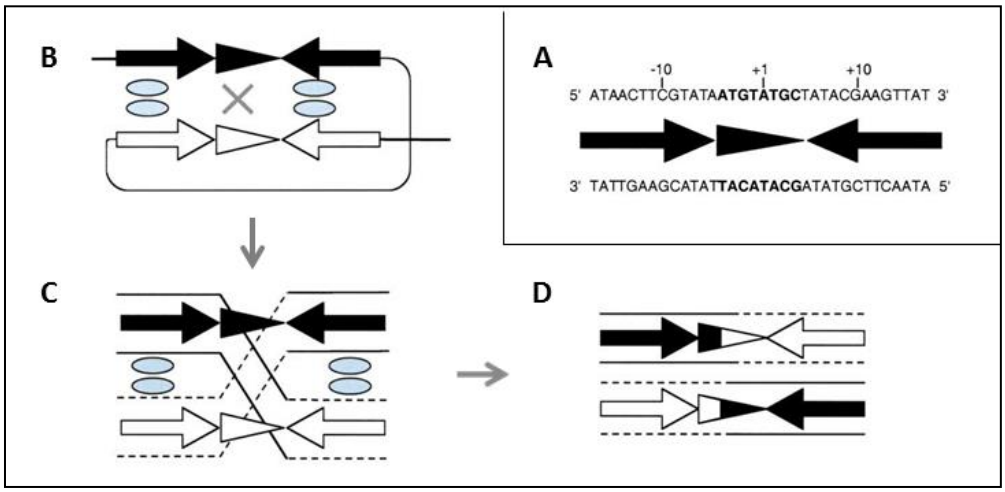


Figure 9: Schematic depiction of the Cre-mediated recombination mechanism. **A)** 34 bp loxP site consisting of an 8 bp asymmetrical core (arrowhead) flanked by two 13 bp inverted repeats [94] (arrows). **B)** Recombination between two loxP sites (black and white) is initiated by the formation of a synaptic complex of 4 Cre monomers (blue) each binding to one of two inverted repeats of a loxP site. **C)** The loxP asymmetrical core is the site of strand cleavage, strand exchange, and ligation. **D)** The recombination process results in the exchange of DNA sequences distal to loxP sites [95].

Due to its simplicity, the system has become a powerful research tool and is used nowadays to tightly regulate the location and timing of genetic manipulations in a wide variety of organisms. Focusing on protozoan parasites, the system was successfully established in *Toxoplamsa gondii*, *Trypanosoma brucei*, and recently also *Plasmodium falciparum* [96-98].

Specifically, the Cre/loxP system has been used for vector insertion into the *T. gondii* genome [96], and for the removal of drug resistance markers in *T. brucei* and *P. falciparum* [98-100]. An interesting Cre/lox-based approach was applied in *T. gondii* by Koshy and colleagues who developed a system for *in vivo* monitoring of parasite infection in mice [101]. They generated a *T. gondii* cell line which injects a Cre-fusion protein into Cre-reporter mouse cells, resulting in expression of GFP upon parasite invasion. Another application of the Cre/loxP system was the generation of knock outs in *T. brucei* and *T. gondii* [99, 102].

3. Goals of the Thesis

3.1. Project 1: Proteomic analysis of *Giardia lamblia* encystation-specific vesicles (ESVs)

For transmission to a new host, *G. lamblia* depends on the regulated secretion of a protective extracellular composite biopolymer, the cyst wall. In the absence of a conventional Golgi apparatus, specialized secretory organelles termed encystation-specific vesicles (ESVs) are synthesized *de novo* during encystation, dedicated to accumulation, maturation, sorting and regulated secretion of the cyst wall material (CWM). As absolutely essential factors for parasite transmission to a new host, ESVs represent potential drug targets and therefore have been subject of intense research for years. However, the molecular underpinning of ESVs, their neogenesis as well as their functionality is not well understood, and a detailed systematic analysis has been impeded by the scarcity of ESV-specific factors and the challenge to identify latter in the highly reduced and diverged parasite genome.

The aim of this study was the identification of novel ESV-associated factors that may help to characterize the nature of these specialized secretory organelles in more detail.
Using various experimental approaches, we addressed the following questions: **i)** Can we define the ESV organelle proteome by using a novel, flow-cytometry based approach, and identify additional factors that are associated with ESVs? **ii)** What can these data tell us about ESV neogenesis, composition, and functionality? **iii)** Based on the newly gained information, can we establish a more detailed model for ESV neogenesis and homeostasis?

3.2. Project 2: Implementation of the Cre/loxP system in *Giardia lamblia*

The three *Giardia* cyst wall proteins (CWPs 1-3) constitute, besides a β-(1,3)-N-Acetyl-Galactosamine (GalNAc) carbohydrate, the major components of the cyst wall, hence play a critical role for the parasite's survival in the environment prior to infection of a new host. The functional integrity of the water-resistant cyst wall strongly depends on the CWP1-3 post-translational maturation, correct sorting and regulated secretion. A pivotal role in this process is attributed to encystation-specific vesicles (ESVs), specialized secretory organelles that are established exclusively in encysting cells for accumulation, maturation and secretion of the CWPs.
The function of individual CWPs in ESV neogenesis and maturation as well as their respective contribution to the formation of the highly resistant cyst wall is of substantial interest not only due to their key roles in successful fecal-oral parasite transmission, but also for the investigation of minimal composite biopolymers conferring resistance to cells. However, detailed information on the CWP function is scarce, and the application of forward genetic approaches to study gene or protein function in *G. lamblia* is strongly limited by the parasite's tetraploid genome and the paucity of suitable drug-resistance markers to select for transgenic cells. As a consequence, relatively straightforward genetic

manipulations such as gene knock-outs have never been realized in *Giardia* so far. Several alternative techniques aiming for gene knock-down have been applied but only with limited success.
To manipulate the *G. lamblia* genome in a more effective manner, we aimed to established the *E. coli* P1 phage-derived Cre/loxP in this parasite. The simple system is based on the Cre-recombinase that mediates recombination between two identical DNA sequences termed loxP. When oriented in the same direction, this results in the excision of the interjacent DNA sequence.

In a series of experiments we addressed the following questions: **i)** Is the phage P1-derived Cre-recombinase expressed in *Giardia* functional? **ii)** Can we use the system to remove a stably integrated, floxed puromycin resistance cassette and other intervening gene sequences from the *Giardia* genome? **iii)** Going further, can we apply the Cre/loxP system for recycling of a puromycin resistance marker to perform serial transfections of the same cell line?

4. References

1. Dobell, C. The Discovery of the Intestinal Protozoa of Man. *Proc R Soc Med* **13**, 1-15 (1920).
2. Nash, T.E., Herrington, D.A., Losonsky, G.A. & Levine, M.M. Experimental human infections with *Giardia lamblia*. *J Infect Dis* **156**, 974-84 (1987).
3. Ankarklev, J., Jerlstrom-Hultqvist, J., Ringqvist, E., Troell, K. & Svard, S.G. Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nat Rev Microbiol* **8**, 413-22.
4. Fletcher, S.M., Stark, D., Harkness, J. & Ellis, J. Enteric protozoa in the developed world: a public health perspective. *Clin Microbiol Rev* **25**, 420-49.
5. Savioli, L., Smith, H. & Thompson, A. *Giardia* and Cryptosporidium join the 'Neglected Diseases Initiative'. *Trends Parasitol* **22**, 203-8 (2006).
6. Touz, M.C. The Unique Endosomal/Lysosomal System of *Giardia lamblia* (2012).
7. Nash, T.E. Antigenic variation in *Giardia lamblia*. *Exp Parasitol* **68**, 238-41 (1989).
8. Keister, D.B. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans R Soc Trop Med Hyg* **77**, 487-8 (1983).
9. Sogin, M.L. Early evolution and the origin of eukaryotes. *Curr Opin Genet Dev* **1**, 457-63 (1991).
10. Dacks, J.B., Walker, G. & Field, M.C. Implications of the new eukaryotic systematics for parasitologists. *Parasitol Int* **57**, 97-104 (2008).
11. Morrison, H.G. et al. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* **317**, 1921-6 (2007).
12. Simpson, A.G. & Roger, A.J. The real 'kingdoms' of eukaryotes. *Curr Biol* **14**, R693-6 (2004).
13. Adl, S.M. et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* **52**, 399-451 (2005).
14. Adl, S.M. et al. The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**, 429-93 (2005).
15. Zambrano-Villa, S., Rosales-Borjas, D., Carrero, J.C. & Ortiz-Ortiz, L. How protozoan parasites evade the immune response. *Trends Parasitol* **18**, 272-8 (2002).
16. Elde, N.C., Long, M. & Turkewitz, A.P. A role for convergent evolution in the secretory life of cells. *Trends Cell Biol* **17**, 157-64 (2007).
17. Tovar, J. et al. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* **426**, 172-6 (2003).
18. Marti, M. et al. An ancestral secretory apparatus in the protozoan parasite *Giardia* intestinalis. *J Biol Chem* **278**, 24837-48 (2003).
19. Dacks, J.B. & Field, M.C. Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. *J Cell Sci* **120**, 2977-85 (2007).
20. Faso, C. & Hehl, A.B. Membrane trafficking and organelle biogenesis in *Giardia lamblia*: use it or lose it. *Int J Parasitol* **41**, 471-80 (2011).
21. Marti, M. et al. The secretory apparatus of an ancient eukaryote: protein sorting to separate export pathways occurs before formation of transient Golgi-like compartments. *Mol Biol Cell* **14**, 1433-47 (2003).
22. Lujan, H.D. et al. Developmental induction of Golgi structure and function in the primitive eukaryote *Giardia lamblia*. *J Biol Chem* **270**, 4612-8 (1995).
23. Soltys, B.J., Falah, M. & Gupta, R.S. Identification of endoplasmic reticulum in the primitive eukaryote *Giardia lamblia* using cryoelectron microscopy and antibody to Bip. *J Cell Sci* **109 (Pt 7)**, 1909-17 (1996).
24. Samuelson, J. et al. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proc Natl Acad Sci U S A* **102**, 1548-53 (2005).

25. Abodeely, M. et al. A contiguous compartment functions as endoplasmic reticulum and endosome/lysosome in *Giardia lamblia*. *Eukaryot Cell* **8**, 1665-76 (2009).
26. Lanfredi-Rangel, A., Attias, M., de Carvalho, T.M., Kattenbach, W.M. & De Souza, W. The peripheral vesicles of trophozoites of the primitive protozoan *Giardia lamblia* may correspond to early and late endosomes and to lysosomes. *J Struct Biol* **123**, 225-35 (1998).
27. Touz, M.C., Kulakova, L. & Nash, T.E. Adaptor protein complex 1 mediates the transport of lysosomal proteins from a Golgi-like organelle to peripheral vacuoles in the primitive eukaryote *Giardia lamblia*. *Mol Biol Cell* **15**, 3053-60 (2004).
28. Rivero, M.R. et al. Adaptor protein 2 regulates receptor-mediated endocytosis and cyst formation in *Giardia lamblia*. *Biochem J* **428**, 33-45 (2010).
29. Gaechter, V., Schraner, E., Wild, P. & Hehl, A.B. The single dynamin family protein in the primitive protozoan *Giardia lamblia* is essential for stage conversion and endocytic transport. *Traffic* **9**, 57-71 (2008).
30. Feely, D.E. & Dyer, J.K. Localization of acid phosphatase activity in *Giardia lamblia* and *Giardia* muris trophozoites. *J Protozool* **34**, 80-3 (1987).
31. Lindmark, D.G. *Giardia lamblia*: localization of hydrolase activities in lysosome-like organelles of trophozoites. *Exp Parasitol* **65**, 141-7 (1988).
32. McCaffery, J.M., Faubert, G.M. & Gillin, F.D. *Giardia lamblia*: traffic of a trophozoite variant surface protein and a major cyst wall epitope during growth, encystation, and antigenic switching. *Exp Parasitol* **79**, 236-49 (1994).
33. Slavin, I. et al. Dephosphorylation of cyst wall proteins by a secreted lysosomal acid phosphatase is essential for excystation of *Giardia lamblia*. *Mol Biochem Parasitol* **122**, 95-8 (2002).
34. Svard, S.G., Meng, T.C., Hetsko, M.L., McCaffery, J.M. & Gillin, F.D. Differentiation-associated surface antigen variation in the ancient eukaryote *Giardia lamblia*. *Mol Microbiol* **30**, 979-89 (1998).
35. Regoes, A. et al. Protein import, replication, and inheritance of a vestigial mitochondrion. *J Biol Chem* **280**, 30557-63 (2005).
36. Smid, O. et al. Reductive evolution of the mitochondrial processing peptidases of the unicellular parasites *Trichomonas vaginalis* and *Giardia* intestinalis. *PLoS Pathog* **4**, e1000243 (2008).
37. Dagley, M.J. et al. The protein import channel in the outer mitosomal membrane of *Giardia* intestinalis. *Mol Biol Evol* **26**, 1941-7 (2009).
38. Jedelsky, P.L. et al. The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia* intestinalis. *PLoS One* **6**, e17285 (2011).
39. Tachezy, J., Sanchez, L.B. & Muller, M. Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia* intestinalis, as indicated by the phylogeny of IscS. *Mol Biol Evol* **18**, 1919-28 (2001).
40. Rada, P. et al. The monothiol single-domain glutaredoxin is conserved in the highly reduced mitochondria of *Giardia* intestinalis. *Eukaryot Cell* **8**, 1584-91 (2009).
41. Lujan, H.D., Mowatt, M.R., Conrad, J.T., Bowers, B. & Nash, T.E. Identification of a novel *Giardia lamblia* cyst wall protein with leucine-rich repeats. Implications for secretory granule formation and protein assembly into the cyst wall. *J Biol Chem* **270**, 29307-13 (1995).
42. Mowatt, M.R. et al. Developmentally regulated expression of a *Giardia lamblia* cyst wall protein gene. *Mol Microbiol* **15**, 955-63 (1995).
43. Sun, C.H., McCaffery, J.M., Reiner, D.S. & Gillin, F.D. Mining the *Giardia lamblia* genome for new cyst wall proteins. *J Biol Chem* **278**, 21701-8 (2003).
44. Konrad, C., Spycher, C. & Hehl, A.B. Selective condensation drives partitioning and sequential secretion of cyst wall proteins in differentiating *Giardia lamblia*. *PLoS Pathog* **6**, e1000835 (2010).

45. Jarroll, E.L., Manning, P., Lindmark, D.G., Coggins, J.R. & Erlandsen, S.L. *Giardia* cyst wall-specific carbohydrate: evidence for the presence of galactosamine. *Mol Biochem Parasitol* **32**, 121-31 (1989).
46. Gerwig, G.J. et al. The *Giardia* intestinalis filamentous cyst wall contains a novel beta(1-3)-N-acetyl-D-galactosamine polymer: a structural and conformational study. *Glycobiology* **12**, 499-505 (2002).
47. Stefanic, S. et al. Neogenesis and maturation of transient Golgi-like cisternae in a simple eukaryote. *J Cell Sci* **122**, 2846-56 (2009).
48. Faso, C., Konrad, C., Schraner, E.M. & Hehl, A.B. Export of cyst wall material and Golgi organelle neogenesis in *Giardia lamblia* depend on endoplasmic reticulum exit sites. *Cell Microbiol* (2012).
49. Hehl, A.B., Marti, M. & Kohler, P. Stage-specific expression and targeting of cyst wall protein-green fluorescent protein chimeras in *Giardia*. *Mol Biol Cell* **11**, 1789-800 (2000).
50. Touz, M.C., Lujan, H.D., Hayes, S.F. & Nash, T.E. Sorting of encystation-specific cysteine protease to lysosome-like peripheral vacuoles in *Giardia lamblia* requires a conserved tyrosine-based motif. *J Biol Chem* **278**, 6420-6 (2003).
51. Hehl, A.B. & Marti, M. Secretory protein trafficking in *Giardia* intestinalis. *Mol Microbiol* **53**, 19-28 (2004).
52. Nash, T.E. Surface antigenic variation in *Giardia lamblia*. *Mol Microbiol* **45**, 585-90 (2002).
53. Nash, T.E., Lujan, H.T., Mowatt, M.R. & Conrad, J.T. Variant-specific surface protein switching in *Giardia lamblia*. *Infect Immun* **69**, 1922-3 (2001).
54. Prucca, C.G. et al. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* **456**, 750-4 (2008).
55. Papanastasiou, P., Hiltpold, A., Bommeli, C. & Kohler, P. The release of the variant surface protein of *Giardia* to its soluble isoform is mediated by the selective cleavage of the conserved carboxy-terminal domain. *Biochemistry* **35**, 10143-8 (1996).
56. Davids, B.J. et al. A new family of *Giardia*l cysteine-rich non-VSP protein genes and a novel cyst protein. *PLoS One* **1**, e44 (2006).
57. Morf, L. et al. The transcriptional response to encystation stimuli in *Giardia lamblia* is restricted to a small set of genes. *Eukaryot Cell* **9**, 1566-76 (2010).
58. Macechko, P.T., Steimle, P.A., Lindmark, D.G., Erlandsen, S.L. & Jarroll, E.L. Galactosamine-synthesizing enzymes are induced when *Giardia* encyst. *Mol Biochem Parasitol* **56**, 301-9 (1992).
59. Reiner, D.S., McCaffery, M. & Gillin, F.D. Sorting of cyst wall proteins to a regulated secretory pathway during differentiation of the primitive eukaryote, *Giardia lamblia*. *Eur J Cell Biol* **53**, 142-53 (1990).
60. Chiu, P.W., Huang, Y.C., Pan, Y.J., Wang, C.H. & Sun, C.H. A novel family of cyst proteins with epidermal growth factor repeats in *Giardia lamblia*. *PLoS Negl Trop Dis* **4**, e677 (2010).
61. Touz, M.C. et al. The activity of a developmentally regulated cysteine proteinase is required for cyst wall formation in the primitive eukaryote *Giardia lamblia*. *J Biol Chem* **277**, 8474-81 (2002).
62. Stefanic, S., Palm, D., Svard, S.G. & Hehl, A.B. Organelle proteomics reveals cargo maturation mechanisms associated with Golgi-like encystation vesicles in the early-diverged protozoan *Giardia lamblia*. *J Biol Chem* **281**, 7595-604 (2006).
63. Castillo-Romero, A. et al. Rab11 and actin cytoskeleton participate in *Giardia lamblia* encystation, guiding the specific vesicles to the cyst wall. *PLoS Negl Trop Dis* **4**, e697 (2010).
64. Davids, B.J. et al. An atypical proprotein convertase in *Giardia lamblia* differentiation. *Mol Biochem Parasitol* **175**, 169-80 (2011).
65. DuBois, K.N. et al. Identification of the major cysteine protease of *Giardia* and its role in encystation. *J Biol Chem* **283**, 18024-31 (2008).

66. Karr, C.D. & Jarroll, E.L. Cyst wall synthase: N-acetylgalactosaminyltransferase activity is induced to form the novel N-acetylgalactosamine polysaccharide in the *Giardia* cyst wall. *Microbiology* **150**, 1237-43 (2004).
67. Gillin, F.D. & Diamond, L.S. Clonal growth of *Giardia lamblia* trophozoites in a semisolid agarose medium. *J Parasitol* **66**, 350-2 (1980).
68. Adam, R.D. Biology of *Giardia lamblia*. *Clin Microbiol Rev* **14**, 447-75 (2001).
69. Adam, R.D. The *Giardia lamblia* genome. *Int J Parasitol* **30**, 475-84 (2000).
70. Best, A.A., Morrison, H.G., McArthur, A.G., Sogin, M.L. & Olsen, G.J. Evolution of eukaryotic transcription: insights from the genome of *Giardia lamblia*. *Genome Res* **14**, 1537-47 (2004).
71. Hausmann, S. et al. Yeast-like mRNA capping apparatus in *Giardia lamblia*. *J Biol Chem* **280**, 12077-86 (2005).
72. Nixon, J.E. et al. A spliceosomal intron in *Giardia lamblia*. *Proc Natl Acad Sci U S A* **99**, 3701-5 (2002).
73. Russell, A.G., Shutt, T.E., Watkins, R.F. & Gray, M.W. An ancient spliceosomal intron in the ribosomal protein L7a gene (Rpl7a) of *Giardia lamblia*. *BMC Evol Biol* **5**, 45 (2005).
74. Roy, S.W., Hudson, A.J., Joseph, J., Yee, J. & Russell, A.G. Numerous fragmented spliceosomal introns, AT-AC splicing, and an unusual dynein gene expression pathway in *Giardia lamblia*. *Mol Biol Evol* **29**, 43-9 (2012).
75. Nageshan, R.K., Roy, N., Hehl, A.B. & Tatu, U. Post-transcriptional repair of a split heat shock protein 90 gene by mRNA trans-splicing. *J Biol Chem* **286**, 7116-22.
76. Kamikawa, R., Inagaki, Y., Tokoro, M., Roger, A.J. & Hashimoto, T. Split introns in the genome of *Giardia* intestinalis are excised by spliceosome-mediated trans-splicing. *Curr Biol* **21**, 311-5 (2011).
77. Kamikawa, R., Inagaki, Y., Roger, A.J. & Hashimoto, T. Splintrons in *Giardia* intestinalis: Spliceosomal introns in a split form. *Commun Integr Biol* **4**, 454-6 (2011).
78. White, T.C. & Wang, C.C. RNA dependent RNA polymerase activity associated with the double-stranded RNA virus of *Giardia lamblia*. *Nucleic Acids Res* **18**, 553-9 (1990).
79. Macrae, I.J. et al. Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195-8 (2006).
80. Saraiya, A.A. & Wang, C.C. snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog* **4**, e1000224 (2008).
81. Ullu, E., Tschudi, C. & Chakraborty, T. RNA interference in protozoan parasites. *Cell Microbiol* **6**, 509-19 (2004).
82. Kabnick, K.S. & Peattie, D.A. In situ analyses reveal that the two nuclei of *Giardia lamblia* are equivalent. *J Cell Sci* **95 (Pt 3)**, 353-60 (1990).
83. Bernander, R., Palm, J.E. & Svard, S.G. Genome ploidy in different stages of the *Giardia lamblia* life cycle. *Cell Microbiol* **3**, 55-62 (2001).
84. Carpenter, M.L., Assaf, Z.J., Gourguechon, S. & Cande, W.Z. Nuclear inheritance and genetic exchange without meiosis in the binucleate parasite *Giardia* intestinalis. *J Cell Sci* **125**, 2523-32 (2012).
85. Ramesh, M.A., Malik, S.B. & Logsdon, J.M., Jr. A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol* **15**, 185-91 (2005).
86. Birky, C.W., Jr. Sex: is *Giardia* doing it in the dark? *Curr Biol* **15**, R56-8 (2005).
87. Cooper, M.A., Adam, R.D., Worobey, M. & Sterling, C.R. Population genetics provides evidence for recombination in *Giardia*. *Curr Biol* **17**, 1984-8 (2007).
88. Yu, D.C., Wang, A.L. & Wang, C.C. Stable coexpression of a drug-resistance gene and a heterologous gene in an ancient parasitic protozoan *Giardia lamblia*. *Mol Biochem Parasitol* **83**, 81-91 (1996).

89. Sun, C.H., Chou, C.F. & Tai, J.H. Stable DNA transfection of the primitive protozoan pathogen *Giardia lamblia*. *Mol Biochem Parasitol* **92**, 123-32 (1998).
90. Singer, S.M., Yee, J. & Nash, T.E. Episomal and integrated maintenance of foreign DNA in *Giardia lamblia*. *Mol Biochem Parasitol* **92**, 59-69 (1998).
91. Davis-Hayman, S.R. & Nash, T.E. Genetic manipulation of *Giardia lamblia*. *Mol Biochem Parasitol* **122**, 1-7 (2002).
92. Stark, W.M., Boocock, M.R. & Sherratt, D.J. Catalysis by site-specific recombinases. *Trends Genet* **8**, 432-9 (1992).
93. Sternberg, N., Hamilton, D. & Hoess, R. Bacteriophage P1 site-specific recombination. II. Recombination between loxP and the bacterial chromosome. *J Mol Biol* **150**, 487-507 (1981).
94. Hoess, R.H., Ziese, M. & Sternberg, N. P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc Natl Acad Sci U S A* **79**, 3398-402 (1982).
95. Branda, C.S. & Dymecki, S.M. Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* **6**, 7-28 (2004).
96. Brecht, S., Erdhart, H., Soete, M. & Soldati, D. Genome engineering of *Toxoplasma gondii* using the site-specific recombinase Cre. *Gene* **234**, 239-47 (1999).
97. Barrett, B., LaCount, D.J. & Donelson, J.E. *Trypanosoma brucei*: a first-generation CRE-loxP site-specific recombination system. *Exp Parasitol* **106**, 37-44 (2004).
98. O'Neill, M.T., Phuong, T., Healer, J., Richard, D. & Cowman, A.F. Gene deletion from *Plasmodium falciparum* using FLP and Cre recombinases: implications for applied site-specific recombination. *Int J Parasitol* **41**, 117-23 (2011).
99. Scahill, M.D., Pastar, I. & Cross, G.A. CRE recombinase-based positive-negative selection systems for genetic manipulation in *Trypanosoma brucei*. *Mol Biochem Parasitol* **157**, 73-82 (2008).
100. Collins, C.R. et al. Robust inducible Cre recombinase activity in the human malaria parasite *Plasmodium falciparum* enables efficient gene deletion within a single asexual erythrocytic growth cycle. *Mol Microbiol* **88**, 687-701 (2013).
101. Koshy, A.A. et al. *Toxoplasma* secreting Cre recombinase for analysis of host-parasite interactions. *Nat Methods* **7**, 307-9.
102. Andenmatten, N. et al. Conditional genome engineering in *Toxoplasma gondii* uncovers alternative invasion mechanisms. *Nat Methods* **10**, 125-7 (2013).

PART IV MANUSCRIPTS

1. Proteomics of secretory and endocytic organelles in *Giardia lamblia*

This study represents the main part of my PhD work. I started elaborating the strategy and the technological basis for this project in the first year of my PhD and worked 3.5 years intensely on it. The idea of simultaneous sorting of two organelle types and *in silico* filtration of mass spectrometry data was developed together with my direct supervisor Prof. Adrian B. Hehl, who supported me in all steps during this project including manuscript writing.

The great majority of this work including the development of novel protocols, the experimental and bioinformatics work as well as analysis and interpretation of the results was performed by me (see figures 1-8 except for figure 3 left and figure 5), with the partial collaboration of Dr. Cornelia Spycher.

The protocol for flow cytometry-based organelle sorting (FAOS) is at the very limits of technology and was established together with Vinko Tosevski from the Flow Cytometry Facility of University of Zurich. In course of this work I have become an independent power user at the Flow Cytometry Facility. Mass spectrometry analysis was performed in collaboration with Dr. Paolo Nanni from the Functional Genomics Center Zurich (FGCZ). All datasets will be made publicly accessible by incorporation onto the *Giardia*DB (www.*Giardia*DB.org) web platform.

Manuscript draft writing as well as figure assembly was performed almost entirely by me. This manuscript was submitted to PLOSone on the 20th of December 2013 and is currently under review.

Proteomics of Secretory and Endocytic Organelles in *Giardia lamblia*

Petra B. Wampfler[1], Vinko Tosevski[2], Paolo Nanni[3], Cornelia Spycher[1, 4]*, Adrian B. Hehl[1]*

[1]Institute of Parasitology, University of Zurich, Zurich (ZH), Switzerland
[2]Institute of Experimental Immunology, University of Zurich, Zurich (ZH), Switzerland
[3]Functional Genomics Center Zurich, Zurich (ZH), Switzerland
[4]Institute of Parasitology, University of Bern, Bern (BE), Switzerland

Keywords: Giardia, encystation, endocytosis, flow cytometry assisted organelle sorting, organelle proteomics

*Corresponding authors

Correspondence and requests for materials should be addressed to:
Adrian B. Hehl
Laboratory of Molecular Parasitology
Institute of Parasitology
University of Zurich
Winterthurerstrasse 266a
8057 Zürich, Switzerland
adrian.hehl@access.uzh.ch
Tel. +41 44 635 8526

or
Cornelia Spycher
Institute of Parasitology
University of Bern
Länggassstrasse 122
3012 Bern, Switzerland
cornelia.spycher@vetsuisse.unibe.ch
Tel. +41 31 631 2396

Abstract

Giardia lamblia is a flagellated protozoan enteroparasite transmitted as an environmentally resistant cyst. Trophozoites attach to the small intestine of vertebrate hosts and proliferate by binary fission. They access nutrients directly via uptake of bulk fluid phase material into specialized endocytic organelles termed peripheral vesicles (PVs), mainly on the exposed dorsal side. When trophozoites reach the G2/M restriction point in the cell cycle they can begin another round of cell division or encyst if they encounter specific environmental cues. They induce neogenesis of Golgi-like organelles, encystation-specific vesicles (ESVs), for regulated secretion of cyst wall material. PVs and ESVs are highly simplified and thus evolutionary diverged endocytic and exocytic organelle systems with key roles in proliferation and transmission to a new host, respectively. Both organelle systems physically and functionally intersect at the endoplasmic reticulum (ER) which has catabolic as well as anabolic functions. However, the unusually high degree of sequence divergence in Giardia rapidly exhausts phylogenomic strategies to identify and characterize the molecular underpinnings of these streamlined organelles. To define the first proteome of ESVs and PVs we used a novel strategy combining flow cytometry-based organelle sorting with in silico filtration of mass spectrometry data. From the limited size datasets we retrieved many hypothetical but also known organelle-specific factors. In contrast to PVs, ESVs appear to maintain a strong physical and functional link to the ER including recruitment of ribosomes to organelle membranes. Overall the data provide further evidence for the formation of a cyst extracellular matrix with minimal complexity.

Introduction

As the leading cause for protozoal diarrhea worldwide, the small intestinal parasite *Giardia lamblia (syn. G. duodenalis, G. intestinalis)* is an important pathogen of humans and animals causing significant morbidity and economic loss [1]. The *Giardia* life cycle is simple and consists of trophozoites, which multiply by binary fission in the gut of animal and human hosts, and an infectious cyst stage. Trophozoites attach actively to the epithelium of the small intestine and exhibit antigenic variation of variant surface proteins (VSPs) in their protein surface coat [2, 3]. Triggered by environmental cues (e.g. bile concentration, bioavailability of lipids, pH) trophozoites undergo a complex stage-differentiation process and transform to environmentally resistant cyst forms. The complete life-cycle, including cyst formation and excystation, can be reproduced in vitro.

Giardia belongs to the phylum Diplomonadida, unicellular eukaryotes that have undergone considerable reductive evolution resulting in minimization or even loss of most cellular systems such as mitochondria, peroxisomes, a Golgi apparatus, and a classical endo-lysosomal system. Despite this unusual organization 3 giardial organelle systems are clearly discernible: the endplasmic reticulum (ER) which extends bilaterally through the cell body [4], relic mitochondria (mitosomes), localized at the cell center but also dispersed in the cytoplasm [5, 6], and peripheral vesicles (PVs). PVs are ~150 nm compartments with a fixed localization underlying the plasma membrane on the dorsal side and are also present in a specialized region of the ventral disk [7, 8]. PVs have been dubbed endosomal-lysosomal compartments based on localization of hydrolase activity [7, 9-11], their ability to acidify [10, 12, 13], and

to take up exogenous ferritin as well as fluid phase markers [12, 14-16]. While in classical eukaryotic systems endosomes undergo organelle maturation, a similar process is not observed for PVs in Giardia. A selective pathway sorting proteins from the plasma membrane to PVs has been demonstrated [13, 17-19], and there is experimental evidence for a direct but selective connectivity between PVs and the ER [8]. PVs are thought to be the major route of nutrient uptake by the parasite, but the range of their functions, their morphogenesis and propagation remain unclear. Additional open questions concern the exact mechanism of bulk fluid uptake into the organelles and how endocytic cargo is sorted and trafficked to the ER.

Constitutive secretion of giardial proteins does not require a Golgi apparatus. As a consequence, secreted proteins are exported directly from the ER to target organelles such as PVs or the plasma membrane [20]. Organelles with Golgi properties (encystation-specific vesicles, ESVs) are generated *de novo* exclusively for regulated export of the cyst wall biopolymer consisting of three paralogous cyst wall proteins (CWP1-3) [21-23], and a unique β(1-3)-GalNAc homopolymer glycan [24, 25]. ESV formation is induced by COPII-dependent export of cyst wall proteins from ER exit sites [26, 27]. CWPs partition into two biophysically distinct phases before being sorted and secreted sequentially to build the two layers of the composite cyst wall polymer as an extracellular matrix 20-24 h post induction (p.i.) in vitro [28]. Although current data strongly support the hypothesis that ESVs are Golgi-derived organelles, the molecular underpinnings of ESV neogenesis and their identity as post-ER organelles remain controversial.

As with most systems and molecular machineries in diplomonads, PV and ESV organelles can be classified in terms of function but are highly divergent, i.e. there is some experimental evidence for their respective roles in the cell, but the paucity of molecular and morphological landmarks for organelle structure and function has prevented a systematic and detailed characterization. Nevertheless both organelle systems are essential and as such represent potentially vulnerable structures of this highly adapted parasite. The significant reductive evolution and sequence divergence in Giardia also means that strategies using homology-based identification and functional analysis of PV and ESV proteins [29-31] are biased towards few identifiable factors. The aim of this study was to identify novel factors that are specifically associated with these organelles and that may help characterize their nature, range of function, and evolutionary history in the context of the giardial ecological niche. To address these questions we developed a conceptually new approach to generate enriched organelle proteome datasets in two steps: (i) simultaneous flow cytometry-based sorting of a mixed microsome fraction containing green fluorescent ESV organelles with CWP3-GFP in condensed cores and red labeled peripheral vesicles (PV); (ii) mass spectrometry analysis and subtraction of overlapping hits to increase identification of organelle-specific candidates. Detailed analysis of the datasets and localization of selected candidate proteins suggests a close association of ESVs with the ER and no evidence for additional cargo or organelle-specific factors involved in their genesis and maturation. Conversely, although direct connections with the ER have been demonstrated, PVs appear to have a discrete compartment identity with a specific set of organelle proteins.

Results

Giardia has an extensive ER [4] which stretches almost throughout the entire cytoplasm and to the cell periphery. Giardial organelle preparations for proteomic analysis are uniformly contaminated with membrane and membrane associated ER proteins [32, 33]. To overcome this we developed and implemented a new strategy to investigate the proteome of two organelle sets which are known to be in close proximity to and even in direct contact with the ER: ESV organelles [26, 27] and PV organelles [8]. After cell disruption we simultaneously enriched differentially labeled ESV and PV organelles by fluorescence-assisted organelle sorting (FAOS) from a mixed microsome fraction. We reasoned that the simultaneously sorted vesicle fractions would contain comparable amounts of mostly cytoplasmic and ER-derived unspecific proteins, and that this overlap could be subtracted from the respective mass spectrometry (MS) datasets *in silico* to reveal organelle-specific proteins.

Enrichment and separation of fluorescently labeled organelles by flow cytometry

We used flow cytometry to sort differentially labeled ESVs (with CWP3-GFP) and PVs (with Dextran-AlexaFluor647, AF647) simultaneously from a mixed microsome fraction prepared from trophozoites and transgenic encysting cells. Cell labeling, harvesting, and disruption were performed in three completely independent experiments (biological replicates). The sorts were performed on a BD FACS AriaIII flow cytometer using a sort precision mode of 0/32/0 to obtain maximal purity. Three gates were set: a very broad parent gate P3 in the SSC vs. FSC plot that excluded the readily apparent measurement noise, and gates P1 and P2 in a bivariate dot plot to define the GFP-positive and AF647-positive events, respectively (figure 1A). The target events in the mixed microsome fraction were 4.3% GFP-positives and 4.7% AF647-positives, corresponding to ESV and PV organelles, respectively.

Post-sort quality control by flow cytometry showed more than eight-fold increase of both, GFP-positive events (39.6%, figure 1B, top) and AF647-positive events (42%, figure 1B, bottom). In the post sort analysis 10^4 events of the AF647-enriched fraction contained no GFP-positive events, i.e. ESV organelles (figure 1B, bottom, gate P1). Likewise, analysis of 10^4 events of the GFP-enriched fraction contained no AF647-positive events, i.e. PV organelles (figure 1B, top, gate P2), with a scatter profile corresponding to that of the pre-sort sample. The cluster of events extending diagonally and between the GFP and AF647 fraction was apparent in all samples analyzed, both pre- and post-sort. Comparisons of normal buffer preparations (used across all experiments) with more meticulously prepared sample buffers and sheath fluid (manually filtered through liquid filters with a pore size of 220 nm) led us to conclude that these events predominantly represent inorganic or similar particulates of very small size. Taken together, GFP- and AF647-positive events were completely separable, and we achieved a 100% relative enrichment using this approach.

To confirm the separation of labeled organelles, we detected GFP in post-sort precipitates (figure 1C) using Western blot analysis. Using an anti-GFP antibody, a band between 40 and 55 kDa corresponding to the predicted CWP3-GFP fusion protein (53 kDa) was detected in the ESV-enriched fraction but not in the PV-enriched fraction. An additional GFP-signal was detected in the high molecular weight area in the ESV-enriched fraction, presumably corresponding to insoluble CWP3-GFP aggregates from condensed cores of ESVs or covalently linked homo-multimers. Taken together with the data obtained by post sort

flow cytometry, we concluded that the two organelles could be quantitatively separated which was a prerequisite for the subsequent subtractive analysis.

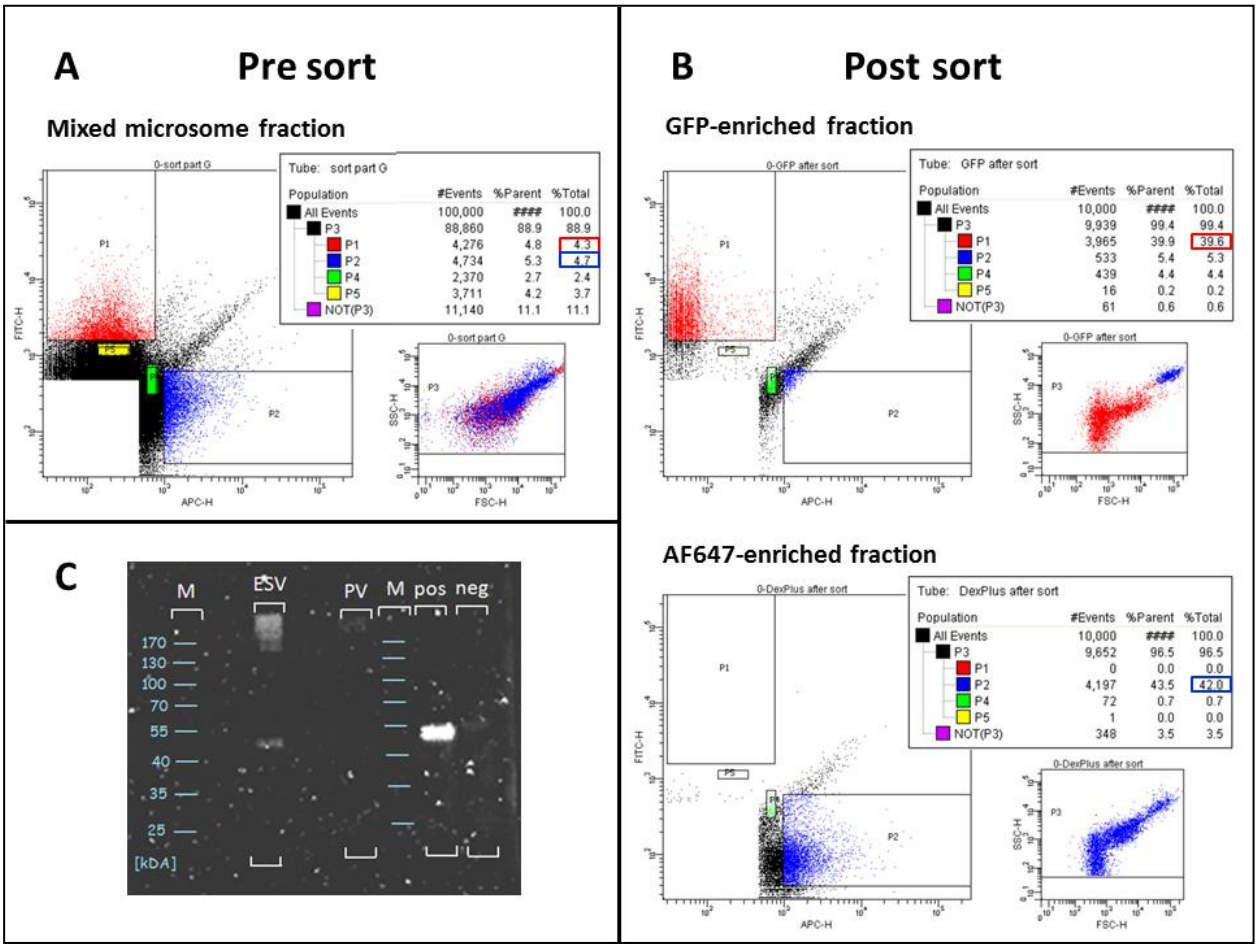


Figure 1: Flow cytometry based organelle sorting. A) Pre-sort flow cytometry analysis: 10e5 events of a mixed microsome fraction containing 4.3% GFP-positive events (ESV organelles) and 4.7% AF647-positive events (PV organelles). SSC/FSC scatter plot: P3, parent gate (organelle population); P1 (red) GFP-positive-, and P2 (blue) AF647-positive events. P4 and P5 (controls) were set to randomly collect additional material for later protein precipitation test runs. **B)** Post sort analysis shows enrichment of GFP-positive events (39.6 %; B, top) and AF647-positive events (42.0%; B, bottom). **C)** Western blot analysis of enriched organelle fractions (post-sort precipitates) detecting GFP shows a distinct band for CWP3-GFP (53 kDA, arrow) in the ESV fraction, but not the PV post sort fraction. *M:* protein size ladder (kDa); *pos, neg:* Positive and negative controls from an ESV-fraction and a non-ESV-fraction (control) of a sucrose density gradient centrifugation experiment performed with 13h encysting Giardia cells expressing CWP3-GFP [28].

Mass spectrometry analysis of organelle-enriched fractions

Sorted organelle-enriched fractions were analyzed by mass spectrometry using a shotgun approach combining 1D-SDS-PAGE and LC ESI-MS/MS. With a Mascot ion cut-off score of 20 for peptide-spectrum matches, a minimum of 2 unique peptides, and a protein probability of 80%, a total of 1281 proteins were identified in the combined triplicate ESV and PV fractions. After subtraction of environmental contaminations, e.g. keratins, 1213 *G. lamblia* proteins remained. A false discovery rate (FDR) of 0.0% for peptides and 0.5% for proteins was calculated by Scaffold. In the unified ESV fractions (E1+E2+E3) a total of 1129 proteins were identified; 750 (66%) thereof were detected in all three samples and 933 (83%) in at least 2 of 3 samples (figure 2A, top). Comparable numbers were found for PV organelles: of 1140 proteins identified in total (P1+P2+P3), 708 (62%) were detected in all three samples and 923 (81%) in at least 2 of 3 samples (figure 2A, bottom). The large overlaps of the P and E datasets, respectively, demonstrated high reproducibility between replicate experiments. For a detailed compilation of identified proteins see table S1.

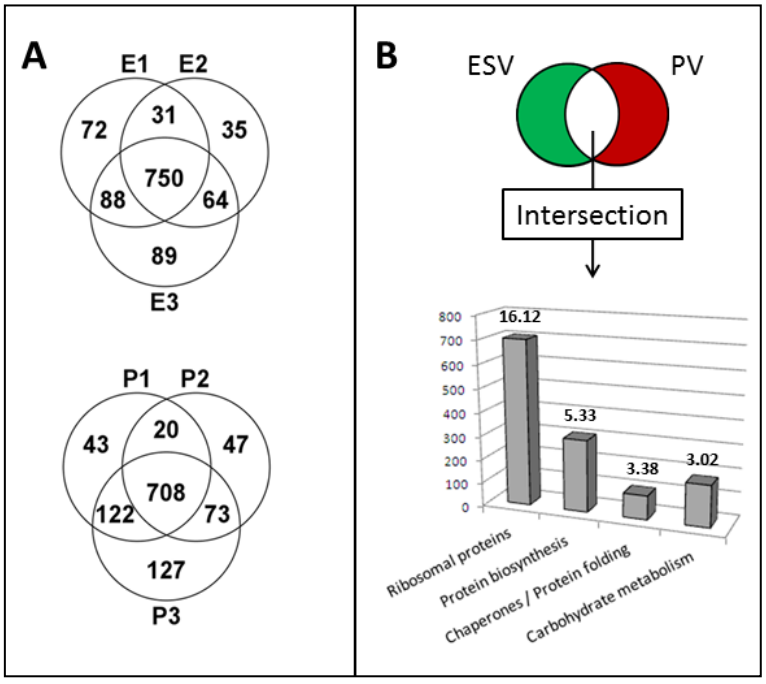


Figure 2: Reproducibility of biological triplicates and annotation clustering of intersection. A) VENN diagrams for ESV (E1, E2, E3) and PV (P1, P2, P3) mass spectrometry datasets (total n=1281). ESV datasets (A, top) n=1129 with overlaps. 750 proteins (66%) were detected in all three datasets. PV datasets (A, bottom) n=1140 with overlaps. 708 proteins (62%) were detected in all three datasets. **B)** Clustering analysis of 1059 proteins defined by the E1-E3 and P1-P3 data intersection (see figure S3) using the DAVID bioinformatics tool [34]. *X-axis:* functional clusters with an enrichment score above 3; *y-axis:* number of proteins; value in bold on top of column: enrichment score. A detailed summary of all 14 clusters can be found in the text S2A and S2B.

Subtractive analysis eliminates many predominantly translation-associated and ER-derived contaminants

Simultaneous sorting of two differentially labeled organelles (ESVs and PVs) from a mixed microsome fraction allows subtracting the unspecific background of unlabeled soluble proteins as well as small cell debris contained within the positively sorted droplets which are generated by the vibrating nozzle of the cell sorter. The premise was that by eliminating all proteins common to the ESV and PV datasets (intersection) the organelle-specificity of each dataset would increase significantly. In particular,

the occurrence of ER-derived contaminants, which have been a severe problem in all previous attempts to enrich Giardia organelles, should be strongly reduced.
The total 1213 hits in all replicate mass spectrometry datasets contained 1059 putative contaminants, defined by the E1-E3 and P1-P3 data intersection, whilst 72 proteins were considered ESV-specific and 82 PV-specific (figure 2B, top). A detailed description of the workflow and the *in silico* identification of the data intersect can be found in the figures S1 and S2. This relatively high ratio of contaminants to organelle-specific proteins was not surprising considering the results of previous cell fractionation experiments and the analysis of organelle fractions by SDS-PAGE in this study (not shown).

Analysis of the eliminated ESV and PV dataset intersection

A more detailed analysis of the two organelle-specific and the large intersecting datasets was performed using the DAVID bioinformatics tool [34]. From 1059 proteins in the dataset intersection we removed an additional 28 with obsolete gene models. Of the 1031 remaining proteins 903 could be assigned to 14 different DAVID clusters (enrichment score >1), while 128 proteins could not be clustered. 4 of the 14 DAVID clusters showed an enrichment score of >3 (figure 2B, bottom). The top ranking clusters were “ribosomal proteins” (enrichment score 16.12, 701 proteins), “protein biosynthesis” (enrichment score 5.33, 305 proteins), “chaperones / protein folding” (enrichment score 3.38, 101 proteins) and “carbohydrate metabolism” (enrichment score 3.02, 176 proteins). A corresponding genome-wide analysis as a reference using a total of 5150 validated genes yielded on average 79 clusters per 3’000 gene models, with the vast majority of enrichment scores <1 and none >2. See also supplementary data (texts S2A and S2B) for a detailed description of all DAVID analyses results. Taken together, this demonstrates the relative enrichment for translation-associated and ER-derived factors in the data intersection, supporting the idea that subtraction of these hits will increase the specificity of organelle datasets. Conversely no obvious clusters were detected in the control datasets representing a genome-wide sampling.

Parsing and manual annotation of ESV and PV organelle specific proteins

The large majority of gene models and protein annotations in the Giardia Genome Database (GiardiaDB) are based on automated predictions. For a manual annotation of organelle-specific datasets we used function and homology prediction programs (PSORTII, TMHMM, SMART, pBLAST and HHPred) to identify putative signal peptides, transmembrane domains, protein functional domains and homologies (table 1 and supplementary table S2). In total we suggest re-annotation of 28 genes (listed in supplementary table S2). For parsing, 17 categories were used (figure 3) based on predicted function or localization according to the criteria described in this study or in previous reports [30].

Table 1: ESV and PV candidate list

ESV candidates (n=72)				PV candidates (n=82)			
Cat.	**GeneID**	**Product description**	**Loc.**	**Cat.**	**GeneID**	**Product description**	**Loc.**
HGS	GI2014	Hypothetical protein		HGS	GI13651	Hypothetical protein	
HGS	GI7350	Hypothetical protein	cy	HGS	GI15918	Hypothetical protein	
HGS	GI9007	Hypothetical protein		HGS	GI16811*	Hypothetical protein	
HGS	GI9157	Hypothetical protein	cy	HGS	GI17468	Hypothetical protein	
HGS	GI10221	Hypothetical protein	er	HGS	GI29119	Hypothetical protein	
HGS	GI10568	Hypothetical protein		HGS	GI6535	Hypothetical protein	
HGS	GI14235	Hypothetical protein		HGS	GI9573	Hypothetical protein	
HGS	GI14458	Hypothetical protein	pe, ex, es	HGS	GI13774	Hypothetical protein	
HGS	GI16522	Hypothetical protein		HGS	GI13783	Hypothetical protein	
HGS	GI22136	Hypothetical protein	er	HGS	GI16926	Hypothetical protein	
HGS	GI24453	Hypothetical protein		HGS	GI17330	Hypothetical protein	
HGS	GI25205	Hypothetical protein	er, es	HGS	GI17571*	Hypothetical protein	
HGS	GI32419	Hypothetical protein	er, ex, es	HGS	GI24451	Hypothetical protein	
HGS	GI87926	Hypothetical protein	vd	HGS	GI3920	Hypothetical protein	
HGS	GI90434	Hypothetical protein		HGS	GI4852	Hypothetical protein	
HGS	GI113722	Hypothetical protein		HGS	GI5890	Hypothetical protein	
HGS	GI137705	Hypothetical protein		HGS	GI6334	Hypothetical protein	
HGS	GI137712	Hypothetical protein		HGS	GI10181	Hypothetical protein	
HNS	GI5568	Hypothetical protein		HGS	GI10524	Hypothetical protein	
HNS	GI8799	Hypothetical protein		HGS	GI10608	Hypothetical protein	
HNS	GI11246	Hypothetical protein		HGS	GI112893	Hypothetical protein	
HNS	GI13262	Hypothetical protein		HGS	GI14971	Hypothetical protein	
HNS	GI14345	Coiled-coil protein		HGS	GI16543	Hypothetical protein	
HNS	GI14877	Hypothetical protein		HGS	GI4018	Hypothetical protein	
HNS	GI15956	WD-40 repeat protein	er	HGS	GI4270	Hypothetical protein	pm, cy, pv?
HNS	GI22806	Hypothetical protein		HGS	GI17347	Hypothetical protein	
HNS	GI23357*	hypothetical protein		HNS	GI91354*	Hypothetical protein	
HNS	GI88581	Synaptic glycoprotein SC2	er	HNS	GI10522	Hypothetical protein	
HNS	GI94654	Hypothetical protein		HNS	GI16237	Hypothetical protein	
MET	GI4059	MTA/SAH-nucleosidase		HNS	GI16367	Hypothetical protein	
MET	GI4507	CTP synthase / UTP-ammonia lyase		HNS	GI16998	Hypothetical protein	
MET	GI6757*	Isochorismatase		HNS	GI7778	Hypothetical protein	
PRO	GI7896	26S proteasome non-ATPase regulatory subunit 7		R/T	GI11287	Ribosomal protein L7Ae	
PRO	GI16823	Non ATPase subunit MPR1 of 26S proteasome		R/T	GI14869	Ribosomal protein L24	
NUC	GI7474	DNA-directed RNA polymerase RPB3		SUG	GI104031	Glycogen synthase, putative	
R/T	GI9899*	putative eIF2A		NUA	GI16328*	pseudouridine synthase	
R/T	GI10341*	putative 50S ribosomal protein L1		MOT	GI13825	Kinesin-1	
R/T	GI10780	Ribosomal protein S27		TPO	GI40224	MDR protein-like protein	er, pe
R/T	GI11319	U3 snRNP IMP3, putative		ER	GI14856	SRP receptor	
R/T	GI11755*	30S ribosomal protein S8E		MIS	GI3095*	Cyclin_N domain containing protein	
R/T	GI15156	SRP GTPase	cy, er, es	MIS	GI8394*	RdX-domain containing protein	

Table 1: ESV and PV candidate list

ESV candidates (n=72)				PV candidates (n=82)			
Cat.	**GeneID**	**Product description**	**Loc.**	**Cat.**	**GeneID**	**Product description**	**Loc.**
R/T	Gl15546*	putative eIF3		MIS	Gl32697*	methyltransferase-domain containing protein	
R/T	Gl40521*	putative eIF2A		MIS	Gl24979*	thioredoxin domain containing protein	
SUG	Gl8382*	Putative UDP-GlcNAc-4'-epimerase (GALE)	er	MIS	Gl2013	Glutaredoxin-related protein	
SUG	Gl10324	Ribulose-phosphate 3-epimerase		MIS	Gl2933	Programmed cell death protein-like protein	
SUG	Gl11595*	Glycosyl transferase family 8 protein	cy	MIS	Gl5871	Developmentally regulated GTP-binding protein 1	
SUG	Gl15483*	UDP-GlcNAc transporter	er, pe, es	MIS	Gl8559	V-type ATPase, 16 kDa proteolipid subunit	er, pe
NUA	Gl16887	ATP-dependent RNA helicase HAS1, putative		TRA	Gl16521	Alpha-SNAP	pv
NUA	Gl113365	5'-3' exoribonuclease 2		TRA	Gl15104	Sec1, putative	pv
MOT	Gl101138	Dynein heavy chain		TRA	Gl15339	AP complex large chain subunit BetaA	er
MOT	Gl111950	Dynein heavy chain		TRA	Gl96994*	Qa3-SNARE	pv
TPO	Gl11299	Amino acid transporter, putative	er	TRA	Gl15472*	VPS46a	cy, pm? pv?
ER	Gl101339	FKBP-type peptidyl-prolyl cis-trans isomerase		P/K	Gl11554	Kinase, NEK	
MIS	Gl6185*	NTPase		P/K	Gl15935	Kinase, NEK	
MIS	Gl7207*	Calcium-binding pr.	cy	P/K	Gl17231	Kinase, NEK	
MIS	Gl8524*	AAA ATPase		P/K	Gl2661	CdK regulatory subunit	
MIS	Gl9594	Hsp70 binding protein		P/K	Gl17069	Kinase, NEK	
MIS	Gl14277	N-acetyltransferase-like protein		P/K	Gl2053	Ser/Thr protein phosphatase 4	
MIS	Gl14604*	cytosolic Fe-S cluster assembling factor NBP35		P/K	Gl17556	Kinase, CAMK CAMKL	
MIS	Gl32838*	Nitrogen fixation protein NifU		P/K	Gl9894	Protein phosphatase 2A regulatory subunit putative	
MIS	Gl113143	Lipopolysaccharide-responsive and beige-like anchor protein		F/C	Gl11164	Protein 21.1	
MIS	Gl9058*	Putative TAP42		F/C	Gl11165	Protein 21.1	
TRA	Gl8049*	Putative importin alpha		F/C	Gl113622	Protein 21.1	
TRA	Gl15204	ERP3		F/C	Gl16532	Protein 21.1	
TRA	Gl17109	Vacuolar protein sorting 11		F/C	Gl17046	Protein 21.1	
TRA	Gl137698	Sec13		F/C	Gl23492	Protein 21.1	
P/K	Gl14545*	putative Ser/Thr-protein phosphatase PP2A		F/C	Gl6744	Centrin	
P/K	Gl14661*	Ser/Thr protein kinase		F/C	Gl95192	Protein 21.1	
P/K	Gl32312	Protein phosphatase 2C		F/C	Gl9750	Intraflagellar transport protein component IFT74/72	
F/C	Gl4026	Alpha-19 giardin		F/C	Gl10038	Alpha-18 giardin	
F/C	Gl15587	Protein 21.1		F/C	Gl10219	Protein 21.1	
F/C	Gl24009	Protein 21.1		F/C	Gl103807	Protein 21.1	
				F/C	Gl13766	Protein 21.1	
				F/C	Gl14926	Protein 21.1	
				F/C	Gl16220	Protein 21.1	
				F/C	Gl14745	Protein 21.1	
				F/C	Gl15428	IFT complex B	
				VSP	Gl33279	VSP	

Table 1: ESV and PV candidate list							
ESV candidates (n=72)				**PV candidates (n=82)**			
Cat.	**GeneID**	**Product description**	**Loc.**	**Cat.**	**GeneID**	**Product description**	**Loc.**
				VSP	GI13194	VSP AS8	
				HCM	GI15317	HCMP Group 1	
				HCM	GI24880	HCMP Group 2	
				HCM	GI103454	HCMP Group 1	

Table 1: ESV and PV candidate list. 72 ESV and 82 PV candidate lists. The subcellular localization of 16 (ESV) and 8 (PV) candidates determined in this study are indicated. Additional information including GenBank accession numbers can be found in the supplementary table S2. *Cat:* category; *GeneID:* Gene accession number according to the *G. lamblia* genome database; *Loc:* localization determined in this study; *er:* endoplasmic reticulum; *pe:* perinuclear ER; *ex:* ER exit sites; *es:* encystation-specific vesicles; *cy* cytoplasm; *vd:* ventral disc; *pv:* peripheral *vesicles; pm: plasma membrane. HGS: hypothetical protein Giardia-specific; HNS: hypothetical protein not Giardia-specific; MET: metabolism; PRO: proteasome; NUC: nuclei; RT: ribosome translation; SUG: sugar metabolism; NUA: nucleic acid metabolism; MOT: motor proteins; TPO: transporter; ER: endoplasmic reticulum; MIS: miscellaneous; TRA: trafficking; P/K: phosphatases / kinases; F/C: flagella/cytoskeleton; VSP: variant-specific surface protein; HCM: high-cysteine membrane protein, pr: protein; asterisk: re-annotated*.

Figure 3 shows a graphical representation of the parsed ESV proteins in 15 categories with 29 hypothetical proteins remaining. Eleven of those (15.3%) have predicted homologs in other species, whereas 18 (25.0%) are considered Giardia-specific. The 82 PV proteins were parsed into 14 categories, with 32 candidates designated as hypothetical proteins and 26 thereof (31.7%) considered Giardia-specific. Direct comparison of the ESV and PV datasets (figure 3) revealed a similar proportion of proteins with no functional prediction, i.e. hypothetical proteins (40.3% and 39.0%, respectively). Parsing of the data revealed clear differences in categories, with 3 and 2 groups appearing exclusively in the ESV and PV datasets, respectively (figure 3). Within the shared groups, the most striking differences were the overrepresentation of factors involved in protein translation (ESV: 11.1% versus PV: 2.4%) in the ESV dataset, or the enrichment of flagella/cytoskeleton proteins (ESV: 4.2% versus PVs: 20.7%) in the PV dataset.

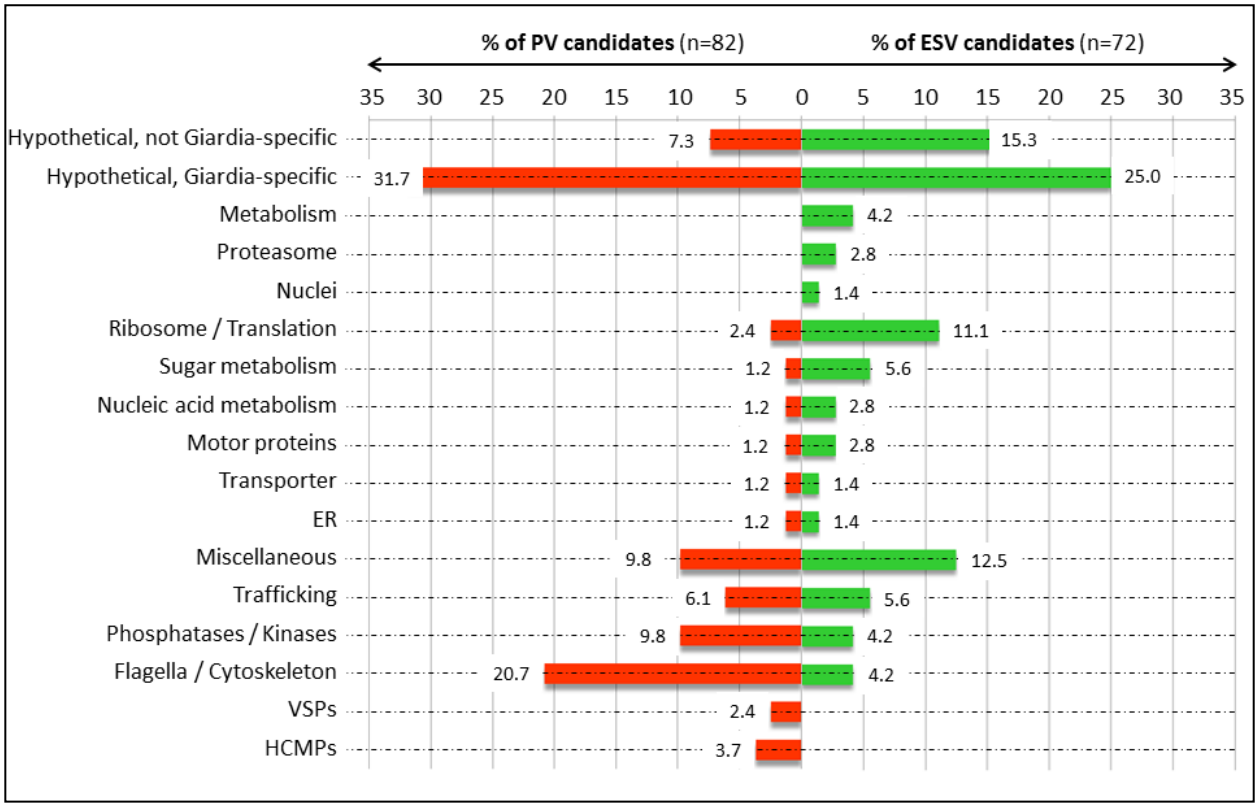


Figure 3: Categorization of the 72 ESV and 82 PV candidates. 17 categories based on predicted function were used. The percentage of candidates in each group is indicated at the end of the columns. ESV candidates (green) and PV candidates (red) fall into 15 and 14 groups, respectively. *ER*: endoplasmic reticulum; *VSPs*: variant surface proteins; *HCMPs*: high cysteine membrane proteins. Detailed information on the candidates in each category can be found in the table S2.

Detection of known organelle proteins in ESV and PV datasets

Fewer than 20 factors are known to associate to ESVs at any given point during ESV neogenesis and maturation [10, 16, 26, 29, 32, 35-39]. Only 3 localize exclusively to ESVs at 13h post induction (p.i.): CWP1 and the large fragment of the proteolytically processed CWP2 in the fluid phase fraction and CWP3 together with the small fragment of CWP2 in the condensed core [28]. CWP-derived tryptic peptides are detected very inefficiently in MS most likely due to the well-documented extensive intra- and intermolecular cross-linking by disulfide and isopeptide bonds [40, 41]. The proportion of DTT-resistant high molecular weight complexes of CWP3-GFP can be estimated in the Western blot in figure 1C (ESV). Nevertheless, peptides derived from the organelle marker CWP3-GFP or endogenous CWP3 were detected exclusively in ESV-enriched samples by MS (average quantitative value of 2.7) albeit only at a stringency of 1 unique peptide and 50% protein probability. GFP-derived peptides were detected in all 3 ESV-enriched samples with an average quantitative value of 1.7.

Of the 72 proteins in the ESV dataset, 4 are either known or predicted to associate to ESV organelles. In particular, we identified two COPII-coat components Sec13 (Gl137698) and Erp3 (Gl15204) (trafficking proteins, figure 3, table 1). The association of other COPII-components with ESVs, such as GlSar1 and GlSec31, during early development was observed previously [26, 29]. Additional ESV-associated factors are represented by 2 proteasome proteins (figure 3, table 1). Proteasome complexes are recruited to ESV membranes during early encystation, perhaps in connection with post-ER quality control and associated degradation processes [32]. It is important to note that with the exception of CWP1-3, all other previously described ESV-associated factors are also expressed in trophozoites. Since their expression is not strictly stage-specific and they are recruited to ESV organelles from other subcellular localizations during encystation, their absence in the ESV-specific dataset but detection in the data intersection instead is not surprising.

The 15 PV proteins which have been identified in trophozoites and encysting cells thus far include SNAREs, components of the clathrin/AP-mediated trafficking machinery, Rab11, an acidic phosphatase and an encystation-specific protease termed ESCP [13, 16-18, 29, 42, 43]. Since most of these PV-associated proteins are involved in vesicle trafficking, they have secondary localizations, i.e. the cytoplasm, the plasma membrane, the ER or ESVs. Not surprisingly, we recovered the majority of these trafficking proteins in the data intersect. However, the Qa3-SNARE homolog syntaxin1 (Gl96994) [29, 43] was specifically included in the PV dataset. In addition to this membrane trafficking factor we recovered 3 previously described PV-associated proteins including two VSPs (Gl33279, Gl13194) [44, 45] and, when lowering the stringency to 1 unique peptide, the encystation-specific cysteine protease ESCP (Gl14566) [13].

Taken together, we detected 5 known ESV-associated proteins, including our organelle marker CWP3-GFP, in the ESV dataset, whereas 4 known PV-associated factors were contained in the PV dataset. Because of their extensive cross-linking and post translational modifications CWP1 and CWP2, comprising the fluid component of the CWM in ESVs at 13 hours p.i., were detected only with relaxed stringency.

Subcellular localization of candidate PV proteins

As a first partial validation of the PV-derived dataset and to determine whether novel PV proteins are contained within the dataset, we expressed 8 proteins as C-terminally HA-tagged variants in *G. lamblia* (table 1, figure 4b). In addition to proteins involved in membrane traffic, we focused on additional predicted transmembrane or membrane interacting proteins. We localized 3 tagged candidates involved in SNARE-mediated membrane fusion: (i) a Qa3-SNARE (GI96994) homologous to syntaxin 1a [29, 43] (ii) a putative Sec1 protein (GI15104) which forms heterodimers with syntaxin 1a [46] and (iii) a putative alpha-SNAP (GI16521). All 3 reporters localized predominantly to the cortical area of trophozoites, consistent with an association to PVs (figure 5A-C). Our localization data of the Giardia Qa3-SNARE are in agreement with previous studies localizing this protein to PVs [43].

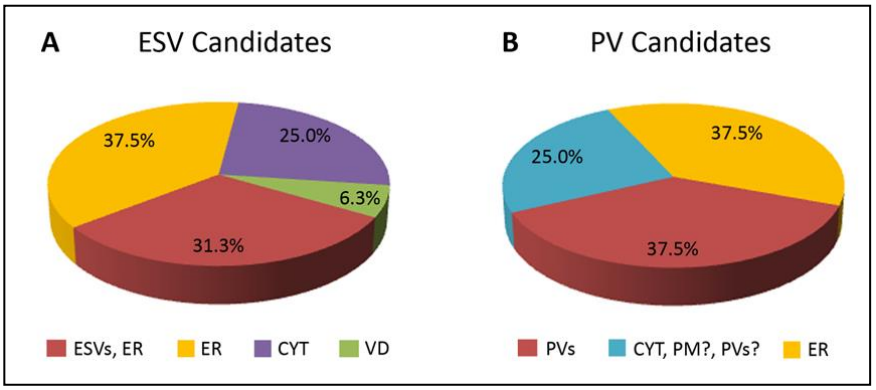


Figure 4: Subcellular localization of selected ESV and PV candidates.
16 (ESV candidates) and 8 (PV candidates) HA-tagged variants were localized by immunofluorescence analysis in transgenic cells. **A)** Subcellular localization of 16 ESV candidates in trophozoites at 13 hours p.i. **B)** Subcellular localization of 8 PV candidates. Detailed information on the respective candidates can be found in table 1 and supplementary table S2. *ER:* endoplasmic reticulum; *ESVs:* encystation-specific vesicles; *CYT*: cytoplasm; *VD*: ventral disc; *PVs:* peripheral vesicles; *PM;* plasma membrane.

We localized two proteins with a predicted function in endocytic/endosomal transport: the adaptor protein (AP) large chain subunit BetaA (GI15339) and a VPS46a homolog (GI15472), a putative component of a giardial ESCRTIII complex. Localization of the tagged AP large chain subunit revealed a distribution consistent with the Giardia ER (figure 5D). This result is partially consistent with earlier studies detecting an AP1-subunit in PVs and the ER, and the AP2-subunit u2 in PVs and the plasma membrane [17, 18]. The tagged giardial VPS46a homolog was detected in the cytoplasm, and probably localizes also to the plasma membrane and/or to PVs (figure 5E).
Subcellular localization studies on the vacuolar ATP synthase subunit (GI8559, figure 5F) and the MDR-like protein (GI40224, figure 5G) showed perinuclear staining and distribution typical of the Giardia ER. One hypothetical protein (GI4270, figure 5H) was detected in vesicle-like structures at the plasma membrane and in the cytoplasm. Whether these structures correspond indeed to PVs will require further investigation.

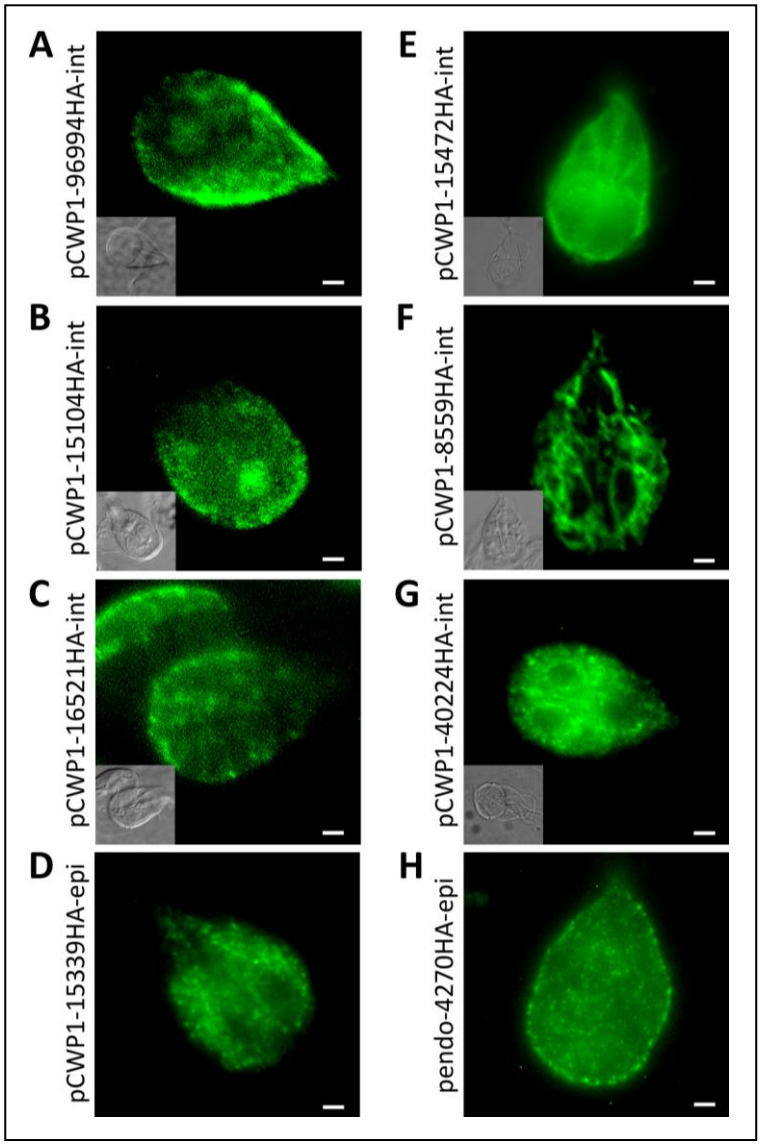


Figure 5: Subcellular localization of 8 selected PV candidates.

Representative localization of C-terminally HA-tagged variants after constitutive expression in trophozoites (H) or 8-12 hours induced expression (A-G). **A, B, C)** Gl96994HA (Syntaxin 1a homolog), Gl15104HA (Sec1 homolog), and Gl16521HA (putative alpha-SNAP) localize to PVs. **D)** Gl15339HA (beta adaptin) localizes to the ER. **E)** Gl15472HA (VPS46 homolog) localizes to the cytoplasm and shows intense signal near the cell surface. **F, G)** Gl8559HA (v-type ATPase) and Gl40224HA (MDR-like protein) localize to the perinuclear region and the ER in trophozoites. **H)** Gl4270HA (hypothetical protein) localizes to vesicle-like structures at the plasma membrane and in the cytoplasm. Localization to PVs cannot be excluded. *Antibodies:* anti-HA high affinity from rat, Alexa488-conjugated goat anti-rat (green), alternatively rat anti-HA-FITC (green). *pCWP1:* inducible CWP1 promoter; *pendo:* endogenous promoter; *int:* stable integration into the genome; *epi:* episomal maintenance of the plasmid. Scale bar: 1.5 µm.

Subcellular localization of candidate ESV proteins

To validate the ESV-organelle dataset, 16 proteins were chosen for ectopic expression as C-terminally HA-tagged variants in *G. lamblia* (table 1, figure 4a) based on the following criteria: (a) Giardia-specific hypothetical proteins since they are unique to Giardia and might associate with ESVs; (b) proteins whose mRNAs were significantly upregulated during encystation [47], representing putative key factors in differentiation; and (c) proteins with predicted transmembrane domains and/or signal peptides, as they may be trafficked to ESVs. In addition, we wanted to localize proteins involved in (d) sugar metabolism or (e) sugar transport as these may represent key factors required for cyst wall glycan synthesis. Another category (f) was composed of predicted transporters which may be involved in direct import of substrates into ESV organelles from the cytoplasm. We also determined the subcellular localization of (g) a newly identified calcium-binding protein. Finally, we selected a signal recognition particle (h) to test for recruitment of complexes for co-translational insertion of proteins across the ER and ESV membranes, as suggested by the electron microscopy data (figure 8).

Subcellular localization studies of 16 ESV candidate proteins revealed 11 candidates which had a distribution consistent with ER localization; 6 were detected exclusively in the ER while 5 also localized to ESVs. Four tagged candidates showed cytoplasmic localization and one candidate localized to the ventral disc.

A) Novel ESV proteins

Only two ER-resident proteins are known to reach ESVs without being secreted to the surface of the cell: Hsp70/BiP, which cycles between the ER and ESVs [32] and a subtilisin-like proprotein convertase termed gSPC [39]. Here we identify 3 additional hypothetical proteins that localize to the ER and ESVs during encystation: **(i)** At 13 h p.i., the 202 amino acid-long HA-tagged GI14458 product was detected in the perinuclear ER, in ER-associated punctate structures reminiscent of ER exit sites, and overlapping with emerging ESVs in the cytoplasm (figure 6B). Transgenic cells expressing GI14458-HA showed a significant delay or, in most cells, a block of CWP1 export from the ER and a reduction in the number of ESVs compared with wild type cells (figure 6A, B) or transgenic cells expressing unrelated HA-tagged products (data not shown). **(ii)** ORF GI32419 is stage-specifically upregulated [48] and codes for a 564 amino acid protein of unknown function (table S2) whose HA-tagged variant localizes in the ER, including structures reminiscent of ER exit sites [27] as well as with maturing ESVs (figure 6C, D). In cells with mature ESVs containing condensed cores the protein localized to ER membranes, in particular to those adjacent to ESVs (figure 6C-E). **(iii)** The predicted GI25205 product is a hypothetical Giardia-specific multipass membrane protein of 1246 amino acids with 14 hydrophobic domains. An epitope-tagged variant localized to the ER and in many cases also to morphologically normal ESVs at 13 hours p.i. (figure 6F).

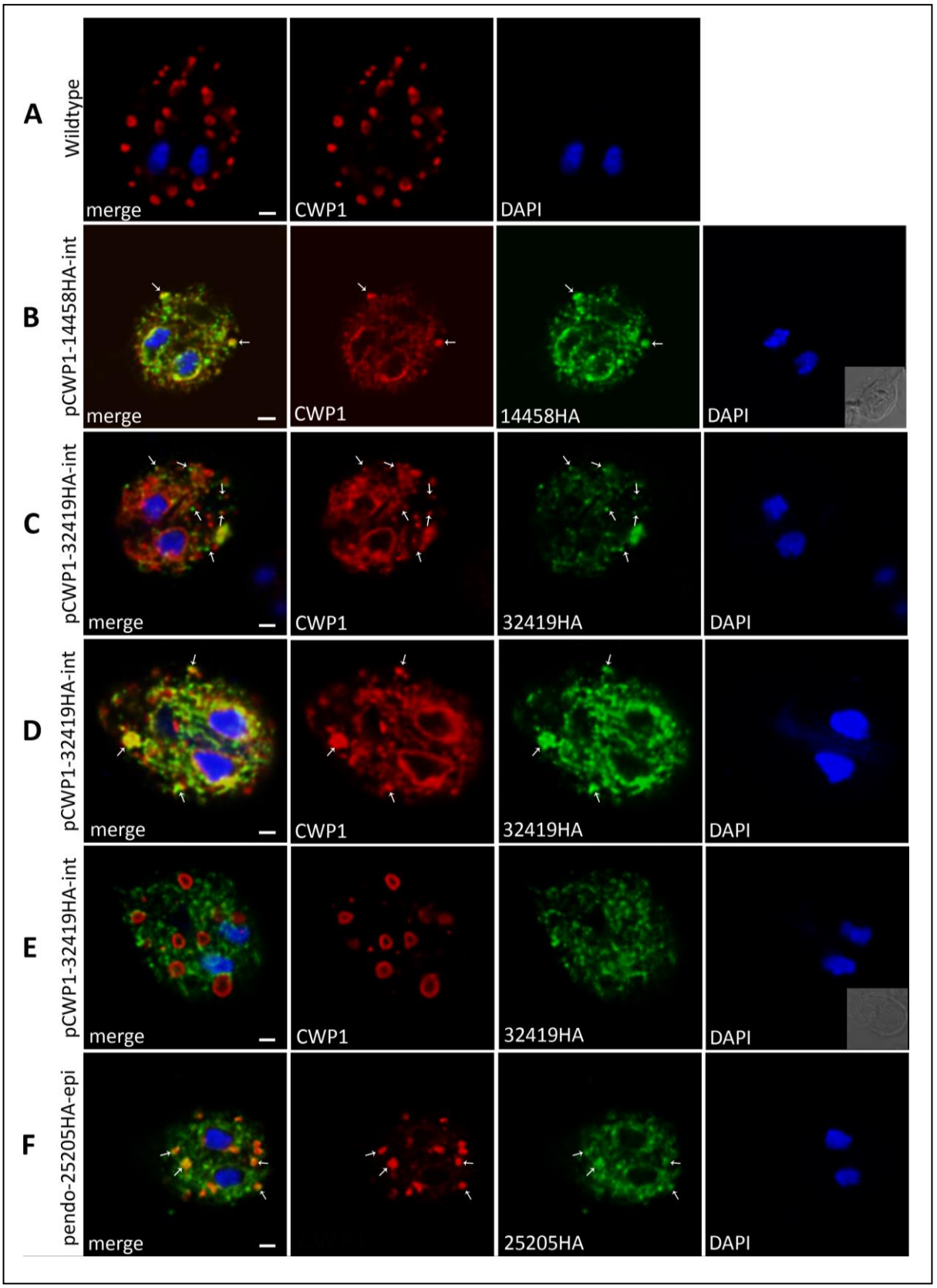

A
Wildtype
merge
CWP1
DAPI
B
pCWP1-14458HA-int
merge
CWP1
14458HA
DAPI
C
pCWP1-32419HA-int
merge
CWP1
32419HA
DAPI
D
pCWP1-32419HA-int
merge
CWP1
32419HA
DAPI
E
pCWP1-32419HA-int
merge
CWP1
32419HA
DAPI
F
pendo-25205HA-epi
merge
25205HA
DAPI

Figure 6: Subcellular localization of three ESV candidates: Gl14458HA, Gl32419HA, and Gl25205HA
Representative localization of C-terminally HA-tagged variants after inducible (Gl14458, Gl32419) or constitutive expression (Gl25205) by confocal microscopy at 13h p.i.. Anti-CWP1 was used to detect ESV organelles. **A)** Wild type (WB) encysting trophozoite at 13h p.i.: CWP1 localized to "doughnut-shaped" ESVs which is typical for this time point [28]. No CWP1-signal in the perinuclear ER was visible. **B)** Gl14458HA at 13h p.i.: Gl14458HA was detected primarily in the perinuclear ER and overlaps with CWP1 in ESVs. Note the retention of CWP1 in the perinuclear ER and reduction of ESV numbers. In fact, the majority of induced cells in the population did not produce ESVs at all. **C, D, E)** Representative subcellular localizations of Gl32419-HA at 13h p.i. **C)** Partial signal overlap of Gl32419-HA (green) with CWP1 (red) in knob-like structures, reminiscent of ESV neogenesis at ER exit sites [27]. Alternatively, Gl32419-HA localized to the ER **D)** co-localizing with CWP1 primarily in the perinuclear ER and in ESVs suggesting delayed export of CWM to ESVs. **E)** In cells with canonical mature ESVs no signal overlap of Gl32419-HA and CWP1 was observed. **F)** HA-tagged Gl25205HA at 13h p.i. was detected in the ER and in ESVs with occasional signal overlap with CWP1. *Antibodies:* anti-HA high affinity from rat, Alexa488-conjugated goat anti-rat (green), and Texas red-conjugated anti-CWP1 (red). Nuclei were labeled with DAPI. *pCWP1:* inducible CWP1 promoter; *pendo:* endogenous promoter; *int:* stable integration into the genome; *epi:* episomal maintenance of the plasmid, *WB:* wildtype. Scale bar: 1.5 µm.

B) Factors involved in CW glycan synthesis

ORF Gl15483 codes for a UDP-GlcNAc sugar transporter [49] with a reported localization at the perinuclear ER and at peripheral vesicles distinct from PVs [50]. This observation is only partially consistent with our localization of an HA-tagged variant in the ER but not to other organelles in encysting trophozoites at 13h p.i. (figure 7A, 7B). In dual labeling experiments, signal overlap of 15483-HA and CWP1 in the perinuclear ER and in areas corresponding to punctate peripheral ER regions or emerging ESVs was observed (figure 7A, arrows). When mature ESVs were present, the overlap of the two proteins was restricted mostly to the perinuclear ER (figure 7B). Notably, Gl15483-HA-expressing cells showed delay of CWP1 export from the ER and accumulation in the perinuclear ER at 13h p.i. (figure 7A, 7B).
ORF Gl8382 codes for a putative UDP-GlcNAc-4'-epimerase (GALE) homolog in Giardia with an N-terminal signal sequence. Manual protein sequence analysis revealed that the giardial protein harbors all conserved motifs required for the enzyme's function [51] (figure S3). An HA-tagged variant localized to the ER (figure 7C). These transgenic cells retained CWP1 in the ER at 13h p.i. and no mature ESVs were formed. The correlation, if any, between this phenotype and the predicted function of the protein, i.e. conversion of UDP-GlcNAc into the UDP-GalNAc monomer of the cyst wall glycan [52] in the ER lumen, remains to be determined.

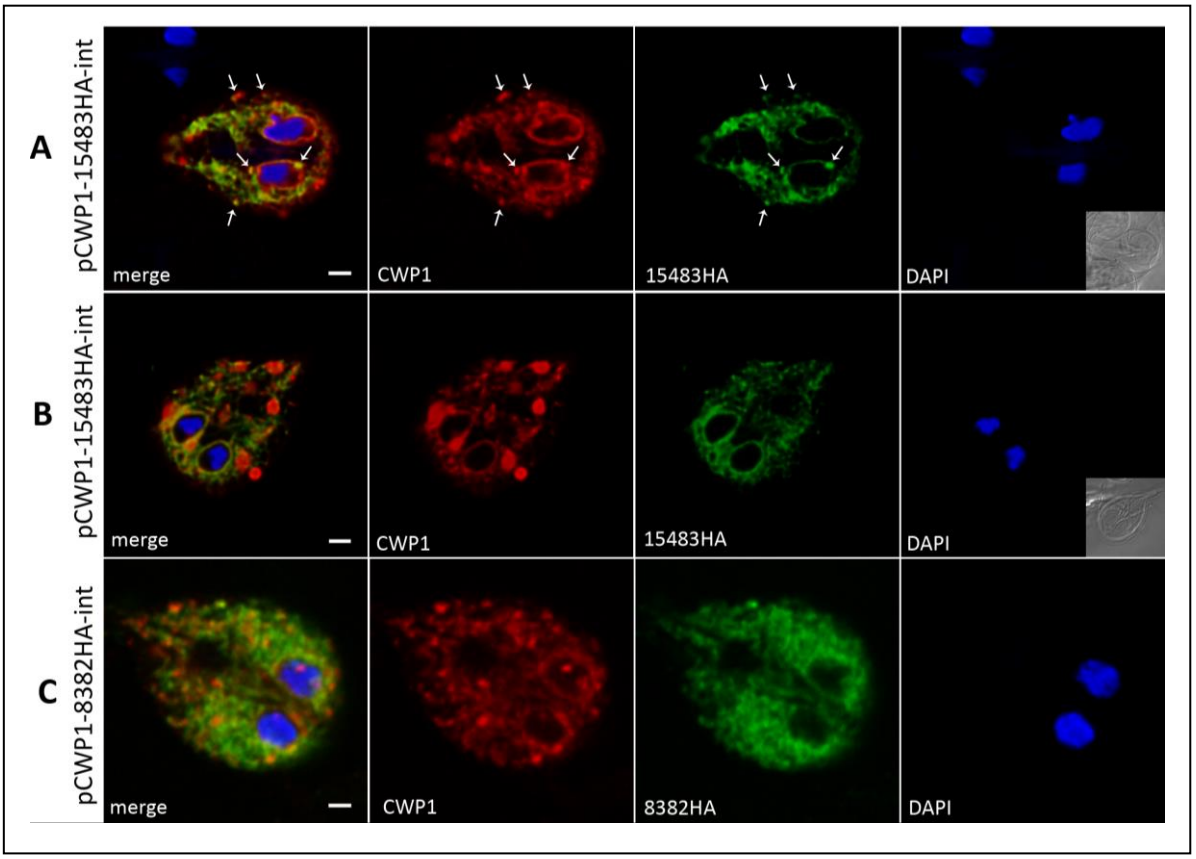


Figure 7: Subcellular localization of the Giardia UDP-GlcNAc transporter (Gl15483) and the putative ER-GALE (Gl8382). Representative subcellular localization of C-terminally HA-tagged variants of UDP-GlcNAc transporter (Gl15483) and the putative UDP-GlcNAc-4'-epimerase (Gl8382) in encysting transgenic cells. **A, B)** At 13h p.i. : the HA-tagged transporter was detected in the ER. Co-localization with CWP1 was observed in the perinuclear ER and in areas corresponding to distinct ER regions or early ESVs (A, arrows). CWP1 is delayed in the perinuclear ER of Gl15483-HA-expressing cells where it overlaps with Gl15483-HA. In mature ESVs, no co-localization of the two proteins was observed. **C)** Localization of the putative UDP-GlcNAc-4'-epimerase (Gl8382-HA) at 13h p.i in the ER together with CWP1 whose export is delayed. *Antibodies:* anti-HA high affinity from rat, Alexa488-conjugated goat anti-rat (green), and Texas red-conjugated anti-CWP1 (red). Nuclei were labeled with DAPI. *pCWP1:* inducible CWP1 promoter; *int:* stable integration into the genome; scale bar: 1.5 µm.

C) Factors for co-translational protein insertion

The panel of candidate ESV proteins contains a substantial number (8, 11.1%) of ribosomal or ribosome-associated proteins (figure 3, table 1) whereas the PV dataset only contained 2 of these potential contaminants. Interestingly, in transmission EM images we found that, in addition to decorating rough ER membranes, ribosomes distinctly associated with ESV membranes (figure 8A, inset arrows). Taken together with the identification of a component of the signal recognition particle, SRP54 (Gl15156) in the ESV dataset, this suggests that ESV cargo proteins could be inserted co-translationally directly into the ESV lumen. Since transgenic cells expressing epitope-tagged ribosomal subunits were not viable, we tested this hypothesis indirectly by expression of an epitope-tagged SRP54 variant (table

1). The tagged product localized in a punctate, distributed pattern and signal overlap with CWP1 was detected predominantly in smaller ESVs (figure 8B). The possibility that proteins, presumably CWPs, are co-translationally inserted directly into ESVs is intriguing but requires additional experimental testing.

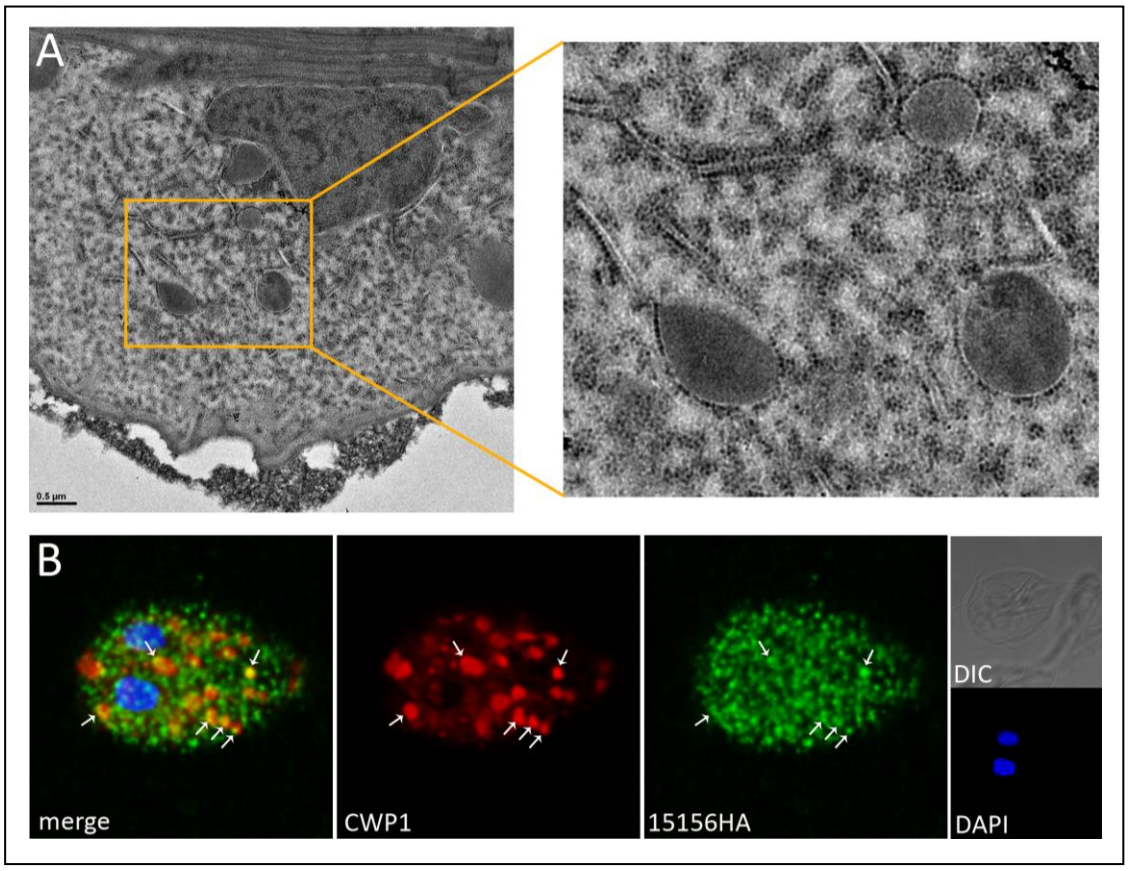


Figure 8: Stage-specific recruitment of ribosomes to ESV membranes

A) Transmission electron microscopy image of ESVs in *G. lamblia* wild type encysting cells at 7 h p.i. (magnified on the right). Recruitment of ribosomes to ER membranes (tubular structures) and to early ESVs (round, electron-dense structures) is observed. Ribosomes are visible as small, round and highly electron-dense structures arrayed along the cytoplasmic side of ESV and ER membranes. **B)** Immunofluorescence analysis of cells expressing a C-terminally HA-tagged signal recognition particle component SRP54 (line pendo-15156HA-epi). The micrograph shows punctuate localization of SRP54-HA and partial overlap with CWP1 accumulation (arrows). Cytoplasmic and ER membrane associated SRP54-HA generates a high background signal, making a detection of the protein at ESV membranes difficult. *Antibodies:* anti-HA high affinity from rat, Alexa488-conjugated goat anti-rat (green), and Texas red-conjugated anti-CWP1 (red). Nuclei were labeled with DAPI. *pCWP1:* inducible CWP1 promoter; *epi:* episomal maintenance of expression vector; scale bar: 1.5 µm.

In summary our localization studies uncovered 5 ESV-associated proteins including 3 Giardia-specific proteins (Gl32419, Gl14458, Gl25205), the signal recognition particle subunit SRP54 (Gl15156), and an UDP-GlcNAc transporter (Gl15483). Six tagged candidates localized exclusively to the ER. Among these were a UDP-GlcNAc-4'-epimerase (GALE) eventually involved in synthesis of the CW glycan, four hypothetical proteins (Gl15956, Gl88581, Gl10221, Gl22136), and an amino acid transporter (Gl11299). Four tagged candidates showed cytoplasmic localization: two Giardia-specific hypothetical proteins (Gl7350, Gl9157), a calcium-binding protein (Gl7207), and a glycosyl transferase family 8 protein (Gl11595). Finally, a Giardia-specific protein of unknown function (Gl87926) was detected at the ventral disc. Taken together, this preliminary analysis suggests that i) abundant novel ESV cargo proteins are not present in mature ESVs, and ii) while large amounts of ER-derived contaminants can be eliminated by subtraction, many ER proteins remain in the dataset. The most likely explanation for this is the intimate physical contact and direct connections between the ER and ESVs as illustrated in figure 8A, which survive cell disruption and organelle preparation.

Discussion

ESV organelles are inducible Golgi-like membrane compartments for accumulation, processing, sorting, and export of the Giardia cyst wall material during differentiation of trophozoites into cysts. Their *de novo* genesis and maturation to secretion-competent organelles is only partially understood: Fewer than 20 ESV-associated factors (among them the 3 CWPs) have been identified or characterized [10, 16, 26, 29, 32, 35-39]. However, no defining ESV-specific, peripherally associated or membrane-bound factor has been identified. Previous attempts to generate an ESV proteome using cell fractionation and density gradient centrifugation yielded datasets which revealed additional ESV factors but also contained very high levels of contaminating proteins [32] (Hehl, unpublished). Here, we tested a conceptually new approach to generate highly enriched organelle proteome datasets for ESVs and PVs resulting in identification of a candidate set of 72 ESV- and 82 PV-associated proteins.

The ER and ESVs maintain a broad range of interactions

Partial validation of the "ESV-specific" protein dataset (72 candidates) by subcellular localization studies of 16 selected ESV candidate proteins revealed 5 proteins which localized to ESVs, but also many ER proteins: 11 candidates had a distribution consistent with ER localization; 6 thereof were detected exclusively in the ER. This suggests that elements of ER organelles remain physically linked to and are co-sorted with labeled ESVs, but not PVs.

ESVs are closely associated with the ER during their neogenesis and nucleated from ERES in a COPII-dependent process [26, 27]. Although light microscopy data suggests that ESVs become physically distinct from the ER after neogenesis [27, 28] previously published EM data [20, 42, 53] and data in the present study (figure 8A) clearly show membrane continuities between the organelle systems. The imaging data is supported by evidence for cycling of ER-resident proteins such as Hsp70/BiP between the two organelles [32]. An extensive network of tubular membrane connections mediating exchange of CWP1 between maturing ESVs makes a definition of the boundaries for ESVs even more challenging [26]: there is a possibility that this dynamic network is not restricted to ESV organelles but establishes direct

connections with the ER [28]. Further, the recruitment of proteasomes [32] and ribosomes to ER and ESV membranes, as shown also in this study, is an additional common feature of the two organelles. This suggests that basic trafficking-related processes such as co-translational import of secreted proteins, folding, retro-translocation, and associated degradation processes start at the ER level but may extend beyond ESV genesis.

Consistent with the premise of subtractive elimination of ER proteins from organelle-specific factors, the MS data intersect was highly enriched in abundant ER-resident proteins, e.g. PDI1-5, Hsp70/Bip and Hsp90/Grp94. On the other hand, many proteins in the ESV dataset showed a typical ER distribution, although 5 proteins also localized to ESVs. A likely explanation is that many tested candidates are present in ER subdomains closely associated or directly connected to ESVs. When encysting Giardia cells are subjected to cell disruption by sonication or other tested methods such as nitrogen cavitation (not shown), these ER domains remain physically linked to ESVs and are enriched accordingly in the sorting process.

Despite elimination of ~90% of all hits in the dataset intersection, which was enriched for known ER proteins, the ESV dataset still contained a large number of proteins localizing to the ER. There are several possible explanations for this surprising finding, the most likely being the intimate association between the ER and ESVs as discussed above. The physical connections and membrane continuities may be resistant to our cell disruption protocols and may thus limit to extraction of ESVs from this subcellular context. In contrast, ER membranes and PV do not occupy the same cellular space: the ER network extends throughout the cytoplasm but not into the cortical layer just below the plasma membrane into which PVs are embedded [54]. Although the interface between the two organelle systems includes postulated direct contact points [8] the subcellular segregation of the two compartments appears to make separation by cell disruption easier.

We performed a preliminary validation of the “ESV-specific” dataset by localizing 16 tagged factors. The results suggested that the dataset contained many ER proteins rather than novel ESV factors. Based on this we propose two non-mutually exclusive interpretations for the lack of novel ESV-specific proteins in our datasets: i) The organelle enrichment strategy and post-sorting elimination of contaminants worked well for filtering out abundant generic ER proteins, but is less efficient in eliminating minor contaminants and/or proteins that remain attached to the fluorescently labeled organelles. ii) ESVs and the ER are highly distinct with respect to abundant luminal proteins such as CWPs, modifying factors, or chaperones. However, both organelles share most membrane or peripherally associated proteins. All available data on stage-specific gene expression during encystation suggests that only CWP genes are strictly stage-specifically regulated (i.e. completely “off” in trophozoites) whilst the remaining (<20) significantly modulated “encystation” genes are upregulated from a basal level in trophozoites during the first 7 hours p.i. [47]. Since CWPs as cargo proteins are certainly more abundant than ESV membrane or organelle-associated proteins, any novel representatives of the latter are difficult to identify within a still relatively high background of ER proteins. In fact, the apparent lack of abundant novel proteins in ESVs exacerbates the appearance of false-positive ER factors in the organelle-specific dataset.

A proposal for cargo-driven ESV neogenesis

The emerging picture in Giardia encystation is that, aside from synthesis of the bulk cyst wall proteins, only relatively small adjustments of expression in some genes (e.g. enzymes for the synthesis of the cyst wall glycan) are required for encysting cells to produce mature ESVs [47, 55]. This also suggests that morphogenesis of ESV organelles could be driven by accumulation of cargo rather than by specific organelle-associated factors, analogous to the formation of dense core secretory granules (DCSG) in endocrine/neuroendocrine cells [56], or in ciliates [57]. In early electron microscopy studies, ESV formation was described as aggregation of electron-dense material in ER membrane-bounded compartments, followed by growth via direct addition of newly synthesized CWPs until large organelles are formed [21, 53, 58]. A more recent model posits that ESV formation is the result of self-organizing properties, mainly of CWP3, leading to formation of a dense core [28]. The ability of CWP1 and 2 to form highly cross-linked complexes [21, 41], and of CWP1 to bind directly to the GalNAc homopolymers [59], which constitutes 60% of the cyst wall material, provides the prerequisites for distributing the extracellular matrix material evenly on the surface of the parasite before initiating polymerization. The presence of basic and simple machinery for ESV formation is supported by the observation that the expression of CWP1 and CWP2 in human embryonic kidney cells is sufficient to induce accumulation in membrane compartments and secretion of the proteins [60]. This observation is in line with granule formation in non-granule forming cells upon expression of different dense core granule cargo proteins, including pro-vasopressin, chromogranin A or von Willebrand factor [61-63], suggesting that the cargo proteins themselves induce the formation of their own carriers through accumulation. Thus, ESV formation might be driven by progressive accumulation of CWPs by a “sorting by retention” mechanism, while ER-resident proteins such as Hsp70/BiP are removed via low-density vesicles or tubular connections between ER and ESVs as the organelles mature [32]. Taken together, the proteome data support a scenario for ESV formation and maturation which relies strongly on inherent properties of cargo proteins and likely only few and as yet unidentified additional components, rather than on a dedicated ESV-specific, organelle-associated machinery driving morphogenesis.
Analysis of the remaining candidates in the data set might bring to light further proteins localizing to ESV organelles. However, none of these genes appear to be significantly upregulated during encystation [47]. Thus, unless significant translational control comes into play [36] we do not expect any additional highly abundant proteins to be discovered exclusively in ESVs.

Taken together, identification of one or more low abundance protein(s) which could be used as defining factors for ESVs as post ER organelles within the only regulated secretory transport pathway in Giardia, remains a significant challenge.

Ribosomal proteins are enriched in the ESV fraction

A comparison of the 72 ESV-specific and 82 PV-specific candidates revealed a significant enrichment of translation/ribosome proteins in the former. Using transmission electron and immunofluorescence microscopy we found support for recruitment of ribosomes not only to the ER but also to ESV membranes during the differentiation process. However, additional functional verification is

required, in particular to test whether co-translational insertion of proteins directed to the regulated secretory pathway may occur directly into ESVs.

In eukaryotic cells, ribosomes localize to the cytoplasm, the nuclear envelope, and the rough ER, giving the latter its typical appearance in transmission electron microscopy. While an association of ribosomes to Golgi membranes in eukaryotic cells containing a steady state Golgi organelle was not observed, giardial ribosomes can be visualized on ESV membranes. Co-translational insertion of secreted proteins directly across ESV membranes might be a consequence of the requirement for producing large quantities of CWP in the relatively short time when ESVs grow maximally. The process of CWP synthesis, translocation and folding clearly begins at the ER level from which CWPs are exported in a COPII-dependent manner [27]). However, the amount of CWPs detected in the ER drops significantly after establishment of small immature ESVs (figure 5 in [27]). Direct co-translational import of CWPs via ESV membranes is one possible explanation for this observation. Directed translocation of proteins across ESV membranes towards the cytosol as part of a quality control system was inferred from the observation that proteasome complexes were recruited to the vicinity of developing ESV organelles [26]. Pore complexes such as Sec61, for which an alpha and a gamma subunit are annotated in the Giardia genome database, are required for co-translational insertion and are also strongly implicated in retro-translocation to the cytoplasm [64]. In support of co-translational insertion of proteins across ESV membranes, a signal recognition particle component, 54 kDa protein (Gl15156), has been localized partially to ESVs (figure 8B). However, none of these factors are distributed in an organelle-specific manner and there is currently no possibility to design experiments allowing a dissection of the directionality of protein translocation across ESV or ER membranes.

Is the cyst wall sugar monomer UDP-GalNAc synthesized in the ER?

The Giardia cyst wall consists of 3 proteins (CWP1-3) and a β(1-3)-GalNAc homopolymer which makes up about 60% of the cyst wall [24, 25]. While the protein components are trafficked via ESV organelles to the surface of the cell, the place of synthesis, transport to the surface, as well as timing and manner of incorporation of the sugar components into the cyst wall remains largely unknown. The sparse literature on the subject suggests synthesis of the cyst wall monomer UDP-GalNAc from endogenous glucose by a series of stage-specifically regulated, enzymatic reactions [52]. A late step, i.e. conversion of UDP-GlcNAc into UDP-GalNAc, was proposed to be performed by a cytosolic UDP-GlcNAc-4'-epimerase or so called GALE (Gl7982). While some experimental data showed that the enzyme converted UDP-GlcNAc into UDP-GalNAc during encystation, investigation of enzyme kinetics showed that the reverse reaction towards production of UDP-GlcNAc was clearly favored, raising significant doubts about a productive synthesis of UDP-GalNAc in the cytoplasm [65].

In this study, two important proteins potentially involved in this process were identified in the ESV dataset: i) the only nucleotide sugar transporter (Gl15483) identified in the Giardia genome project [49] which specifically transports UDP-GlcNAc from the cytoplasm to the ER lumen [50], and ii) a putative UDP-GlcNAc-4'-epimerase (GALE). We detected epitope-tagged variants of the epimerase in the ER, and the transporter mainly showed distribution in the perinuclear ER and early ESVs, where its signal overlapped with that of CWP1. N-glycosylation of Giardia proteins is restricted to addition of $GlcNAc_{1-2}$ to

asparagine [66]. Consistent with this, the parasite lacks genes required for synthesis of the typical eukaryotic core-oligosaccharide $GlcNAc_2Man_9Glc_3$ and for further N-glycan processing in the ER and Golgi. While the UDP-GlcNAc-transporter Gl15483 in the ER membrane imports UDP-GlcNAc used for N-glycosylation, the presence of an ER-localized UDP-GlcNAc-4'-epimerase converting UDP-GlcNAc into UDP-GalNAc indicates involvement of the putative GALE enzyme (Gl8382) in producing the UDP-GalNAc monomer for the cyst wall glycan in the ER.

Opportunities and technical limitations of dual organelle sorting and in silico processing of mass spectrometry data

In proteomic studies, the purity of the biological sample is of utmost importance for a successful analysis. One of the most crucial steps is subcellular fractionation. Despite considerable efforts to optimize protocols for purification of Giardia organelles, the levels of contaminating proteins from non-target organelles and cellular structures remain high [32, 33]. The most frequently used subcellular fractionation techniques applied in organellar proteomics are density-based gradient centrifugation, affinity-based isolation, free flow electrophoresis, and recently also flow cytometry [67-70]. Fluorescence-based organelle sorting by flow cytometry is challenging because of the small size of organelles which usually results in reduced fluorescence intensity. In the case of Giardia organelles, labeling with a highly expressed luminal GFP-tagged organelle marker (CWP3-GFP) in ESVs or the endocytic uptake of a fluid-phase fluorescent dye by PV organelles created unique opportunities for intense organelle labeling. This was sufficient to clearly detect and enrich the organelles simultaneously by flow cytometry despite their small size and, most importantly, to achieve a 100% relative enrichment (i.e. 100% separation) of ESVs and PVs which was a precondition for the subsequent subtractive approach. Analysis of ESV and PV-enriched samples by shotgun mass spectrometry revealed high reproducibility between 3 independent experiments.

The limited purity of organelle preparations and the high sensitivity of current mass spectrometers require additional measures to address the large quantities of false-positive hits. Researchers have developed different (*in silico*) strategies for the elimination of contaminating proteins from organellar datasets. A proteomic study on Giardia mitochondria-relic organelles (mitosomes) using gradient centrifugation took advantage of the organelle distribution into two neighboring fractions [33]. Using iTRAQ mass spectrometry the relative distribution of mitosomal marker proteins between the two fractions was evaluated, and novel putative proteins were identified based on their similar distribution ratio. Another study focusing on the proteome of Spironucleus hydrogenosomes utilized the distribution of the target organelle into two gradient fractions of distinct densities [71]. After mass spectrometry analysis, putative organelle-specific proteins were identified by their co-purification with organelle marker proteins. Both approaches significantly reduced the incidence of contaminants in the resulting large datasets, but the proportion of organelle-specific proteins remained low.

In our case, simultaneous enrichment of ESV and PV organelles makes subtractive approaches to identify contaminating MS hits *in silico* possible. Using this approach we removed 1059 hits which were common to both organelle fractions, including a large proportion (86%) of ribosome/translation and ER-derived

contaminants which constitute a major challenge in organelle proteomic studies. Accordingly, among the 72 ESV and 82 PV candidates only a single predicted ER protein was identified in each dataset. However, preliminary evaluation revealed a large proportion of hitherto unknown ER proteins.

Two factors may contribute to the low discovery rate of organelle proteins: i) peripherally associated organelle proteins may be partially or completely lost during the purification and sorting process, ii) the subtractive approach to remove contaminating proteins likely also eliminates specific categories of organelle-associated proteins, i.e. those with secondary localizations or with large cytoplasmic pools. Examples are the small GTPases Rab1, Rab11, Arf1 and COPI components which also localize to the ER and other membranes. In addition, some peripherally associated factors are recruited to ESVs only during specific phases of the differentiation process, e.g. Rab1, COPI [26, 29], or members of the SNARE family [43].

In summary, the success of a simultaneous sorting approach strongly depends on how cleanly the differentially labeled organelles can be prepared during cell disruption and how similar their cellular context is. The major limitation in our case is the close association of ESVs but not PVs with the ER subdomains. Since the cellular context of the two analyzed organelles differs greatly, subtractive analysis appears to be more efficient for PV candidates but still led to many false-positive candidates in the ESV dataset. Thus, subtractive analysis of datasets derived from simultaneously sorted organelles is a useful strategy to discover organelle-specific factors, but the degree of success depends strongly on the feasibility clean extraction of target organelles from their subcellular context.

Materials and Methods

Giardia cell culture and in vitro encystation

Giardia lamblia WBC6 (ATCC catalog number 50803) trophozoites were grown under anaerobic conditions in 11 ml culture tubes (Nunc, cat. 156758) or triple flasks (Nunc, cat. 132867) containing TYI-S-33 medium supplemented with 10% fetal bovine serum and bovine bile according to standard protocols [41]. Parasites were harvested by chilling the tubes on ice for 30 minutes (for flasks: 1 hour in ice water) to detach adherent cells and collected by centrifugation (900 x *g*, 10 minutes, 4 °C). Encystation was induced using the two-step method as described previously [47] by cultivating the trophozoites in bile-free medium for 44 hours and thereafter in medium (pH 7.85) containing porcine bile.

Sample preparation for flow cytometry

ESV-organelle staining: CWP3-GFP expressing cells [28] and WB wild type cells (control) were grown in triple flasks (800 ml) and kept in encystation conditions for 13 hours as described above. Cells were harvested and resuspended in 20 ml of encystation medium. To allow oxidative chromophore formation, the cell suspension was dispersed into a 6 well plate (Sigma, cat. Z707759) and exposed to air on ice over night. To complete GFP folding, the cells were harvested from the plate and incubated for 30 minutes in a water bath at 37 °C in anaerobic conditions.

PV-organelle labeling: wild type trophozoites were grown in triple flasks and harvested as described above. Cells were washed twice in 10 ml 1x PBS (900 x *g*, 10 minutes, 4 °C) and resuspended in 500 ul supplemented PBS (5 mM glucose, 5 mM cysteine, 0.1 mM ascorbic acid) containing 4 mg/ml dextran AlexaFluor-647 (Molecular Probes Inc., cat. D22914). Endocytic uptake of the fluorescent dye by PV organelles was achieved at 37 °C for 30 minutes, covered from light.
All samples were washed twice in 10 ml 1x PBS (900 x *g*, 10 minutes, 4 °C) and resuspended in 5 ml 1x PBS. After addition of protease inhibitor cocktail (Calbiochem, cat. 539131) and PMSF (Sigma, cat. P7626), the cells were disrupted by four rounds of mild sonication (Branson Sonifier 250, Branson Ultrasonics Corporation, 4x 60 pulses, duty cycle 20%, output control 1.5) on ice. To remove remaining intact cells and cysts, the cell suspensions were passed through a 5 um filter (MILLEX-SV 5.00 um, Millipore, cat. SLSV25LS).

Flow cytometry-based organelle sorting

Flow cytometry-based sorting of organelles was performed on a BD FACS AriaIII™ cell sorter. For data acquisition and processing, the BD FACSDiva™ software (version 6.1.3) was used. In order to achieve maximal speed the sort was performed with using a nozzle with a 70 micrometer orifice diameter at 4.83 bar sheath pressure. GFP was excited by a 488 nm laser and emission was detected using a 525/45 band pass filter. AlexaFluor647 was excited by a 633 nm laser and emission was detected using a 670/40 band pass filter. Given the small organelle size, considerable proportion of observed events was stemming from particulate and electronic noise. All the fluorescent and light scatter parameters were estimated by the height of the voltage pulse generated by each event. The detection threshold was defined as a logical combination of green (GFP) and red (AF647) signal value using the "OR" functional operator. The organelle populations were defined by a parent gate (P3) based on FSC-H and SSC-H (figure 1). YFP control beads (SHERO™ Fluorescent Nanospheres, Spherotech Inc., cat. FP-0552-2) in the size range of the organelles (400 to 600 nm) were used to estimate the level of particulate and/or electronic noise with selected instrument settings. To select for GFP-positive and AF647-positive events out of the mixed organelle population, gates P1 and P2 were set in a bivariate dot-plot. An unlabeled cell suspension was used as negative control to define the gate positions. Prior to the sort, the two cell lysates were mixed in a ratio of about 1:4 to obtain similar numbers of target events, i.e. GFP-positive and AF647-positive events, in the mixture. To attain maximal purity, a sort precision mode of 0/32/0 was chosen. The sort was performed with an average event rate of 25’000 to 30’000 per second, an average sort rate of 600 events per second and a mean sort efficiency of 80%. Twenty million GFP-positive and AF647-positive events were collected in separate 5 ml polystyrene tubes (BD Biosciences, cat. 352052). For quality control of the sort, the collected material was analyzed by flow cytometry using the same settings as for the sort (figure 1).

Protein precipitation

Protein precipitation was performed using the PRM method [72]. Briefly, PRM reagent was added to the sample in a ratio of 1:4. The samples were vortexed and incubated at 25°C for 25 minutes. Proteins were pelleted at 3800 x *g* for 30 minutes and dissolved in 1 ml of ddH2O. After addition of 250 ul of PRM reagent, incubation and pelleting was repeated one more time. The final pellet was resuspended in 25 ul

of Laemmli buffer containing 0.5 % (v/v) β-mercaptoethanol, incubated for 5 minutes in boiling water, and stored over night at -20 °C.

SDS-PAGE and immunoblot analysis

SDS-PAGE on a 12% polyacrylamide gel and subsequent transfer to a nitrocellulose membrane (Protran, Whatman GmbH, cat. 10401396) was performed according to standard protocols. The following antibodies were used: Anti-GFP from mouse (JL-8, Clontech, cat. 632380; 1:2000) and a horseradish peroxidase-conjugated rabbit anti-mouse IgG (Sigma, cat. A9044, 1:8000). Signal detection was performed using Western Lightning Chemiluminescence Reagent (PerkinElmer Life Sciences, cat. NEL100001EA). Data collection was done in a Multimage Light Cabinet with AlphaEase FC software (Alpha Innotech, San Leonardo, CA, USA).

Mass spectrometry analysis and protein identification

Protein samples were boiled for 5 minutes and centrifuged for 1 minute at 16'100 x *g* at room temperature to pellet insoluble material. The samples were separated by 1D-SDS-PAGE using precast 12% Tris-Glycine gels (Invitrogen, cat. IM6000). 4 to 8 ul of supernatant were loaded, depending on the estimated amount of protein in the respective sample determined in a preceding test run. After staining with Roti Blue (CARL ROTH, cat. A152), each gel band was cut into 21 slices. In-gel digestion of proteins using trypsin and extraction of peptides was performed according to standard protocols. Samples were analyzed on a LTQ-Orbitrap XL mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to an Eksigent-Nano-HPLC system (Eksigent Technologies, Dublin, CA, USA). A detailed description of sample preparation and mass spectrometry analysis can be found in the text S1.
The raw-files from the mass spectrometer were converted into Mascot generic files (mgf) with Mascot Distiller software 2.4.2.0 (Matrix Science Ltd., London, UK). The peak lists were searched using Mascot Server 2.3 against the *G. lamblia* database (http://tinyurl.com/37z5zqp) with a concatenated decoy database supplemented with contaminants, The Arabidopsis Information Resource (TAIR9) protein database and the Swissprot database to increase the database's size. The final database included 79141 entries. The identification results were loaded into Scaffold 3.0 (Proteome Software, Portland, US) and filtered for a minimal mascot score of 20 for peptide probability, a protein probability greater than 80%, and a minimum of 2 unique peptides per protein.

In silico removal and functional annotation clustering of the intersection dataset

The final ESV and PV organelle-specific datasets were compiled by *in silico* identification and removal of the contaminating proteins. Briefly, the ESV-derived and PV-derived mass spectrometry (MS)-datasets from the three independent experiments were intersected separately. Proteins detected in both ESV and PV MS-datasets were considered as "contaminants" and removed, thus generating three independent subtractive lists enriched for putative ESV-specific hits and PV-specific hits, respectively. To enhance the stringency for detection of putative organelle-specific candidates, we accepted only ESV candidates that were detected in at least two more ESV MS-datasets than PV MS-datasets, and vice-versa. A detailed description of the procedure is attached in the supplementary figure S3. The contaminating proteins defined by the data intersection were evaluated and clustered into functional groups using the DAVID

bioinformatics tool (http://david.abcc.ncifcrf.gov/home.jsp) [34]. Since analysis in DAVID is restricted to 3000 genes, clustering of the Giardia genome as a control was performed by the generation of 10 independent lists each containing 3000 randomly selected Giardia genes.

In silico analysis tools

Analysis of primary structure and domain architecture of ESV and PV candidates (i.e., manual annotation) was performed using the following tools and databases: PSORTII (http://psort.hgc.jp/form2.html) for prediction of subcellular localization, TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) for prediction of transmembrane helices, SMART (http://smart.embl-heidelberg.de/) for prediction of patterns and functional domains, pBLAST for protein homology detection (protein blast by NCBI, http://blast.ncbi.nlm.nih.gov/Blast.cgi), HHPred (http://toolkit.tuebingen.mpg.de/hhpred) for protein homology detection based on HMM-HMM comparison, and the Giardia genome database (http://giardiadb.org/giardiadb/) for changes in mRNA expression during the Giardia life cycle. For functional domains predicted by SMART we used an e-value of 10e-5 as cutoff, and for protein homologies predicted by pBLAST we accepted alignment scores above 80. Alignment scores between 50 and 80 were accepted only when the pBLAST predictions were consistent with those of HHPred. The latter was used to make pBLAST more robust; only hits with a probability above 95% were accepted.
Functional annotation clustering of the data intersect was performed using the DAVID bioinformatics tool (http://david.abcc.ncifcrf.gov/home.jsp) [34].

Expression constructs and transfection

For cloning of C-terminally HA-tagged proteins in Giardia, a vector PAC-CHA with additional restriction sites was designed on the basis of the previously described vector pPacV-Integ [26]. Additional restriction sites were inserted via oligonucleotide primers. A detailed vector map can be found in the figure S4. For each gene of interest two expression vectors were constructed, one in which expression of the gene of interest is driven by its own promoter (pendo), and another in which the gene of interest is under the control of the inducible cyst wall protein 1 promoter (pCWP1). GenBank accession numbers and a list of primers used for cloning can be found in tables S2 and S3, respectively. For transfection, 15 ug of plasmid DNA linearized with *Swa*I was electroporated (BIO RAD Gene Pulser, 350V, 960 mF, 800 Ohm). The expression vector is targeted to the *Giardia lamblia* triose phosphate isomerase (Gl-TPI) locus by homologous recombination [73] stable transfectants are selected with the antibiotic puromycin (Sigma, cat. 7699111) at a concentration of 77 uM for 5 days. For episomal maintenance, circular plasmid DNA was electroporated and selected with puromycine.

Immune Fluorescence Assay

Immunofluorescence analysis was performed as described previously [20]. The following antibodies were used in this work: Anti-HA high affinity from rat (Roche diagnostics AG, cat. 11867423001, 1:50), Alexa488-conjugated goat anti-rat (Invitrogen, cat. A11006, dilution 1:200), Texas Red-conjugated anti CWP1 (Waterborne^{TM} Inc., cat. A300TR-R, dilution 1:50). For microscopy cells were embedded in Vectashield (Vector Labs, Inc., cat. H-1200) containing the DNA intercalating agent 4'-6-Diamidino-2-phenylindole (DAPI) for staining of nuclear DNA. Immunofluorescence analysis was performed on the

standard fluorescence microscopes Leica DM IRBE with MetaVue software version 5.0r1, or Nikon Eclipse 80i with Openlab Improvision software 5.5.2 for data collection. WCIF ImageJ was used for image processing. Alternatively, analysis was performed on a Leica SP2 AOBS confocal laser-scanning microscope (Leica Microsystems, Wetzlar, Germany) equipped with a glycerol objective (Leica, HCX PL APO CS 63x 1.3 Corr).

Transmission electron microscopy

Transmission electron microscopy and sample preparation was performed as described previously [29].

Acknowledgements

We thank Therese Michel, Katja Huggel, Vreni Balmer, Dr. Christoph Lippuner, and Dr. Claudia Dumrese for excellent technical assistance. We thank Dr. Carmen Faso and Elisabeth Schraner for help with electron microscopy and Dr. Carmen Faso for reviewing of the manuscript.

References

1 Adam RD (2001) Biology of Giardia lamblia. Clinical microbiology reviews 14: 447-475.
2 Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, et al. (2008) Antigenic variation in Giardia lamblia is regulated by RNA interference. Nature 456: 750-754.
3 Nash TE (1989) Antigenic variation in Giardia lamblia. Experimental parasitology 68: 238-241.
4 Soltys BJ, Falah M, Gupta RS (1996) Identification of endoplasmic reticulum in the primitive eukaryote Giardia lamblia using cryoelectron microscopy and antibody to Bip. Journal of cell science 109 (Pt 7): 1909-1917.
5 Tovar J, Leon-Avila G, Sanchez LB, Sutak R, Tachezy J, et al. (2003) Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature 426: 172-176.
6 Regoes A, Hehl AB (2005) SNAP-tag mediated live cell labeling as an alternative to GFP in anaerobic organisms. Biotechniques 39: 809-810, 812.
7 Feely DE, Dyer JK (1987) Localization of acid phosphatase activity in Giardia lamblia and Giardia muris trophozoites. J Protozool 34: 80-83.
8 Abodeely M, DuBois KN, Hehl A, Stefanic S, Sajid M, et al. (2009) A contiguous compartment functions as endoplasmic reticulum and endosome/lysosome in Giardia lamblia. Eukaryot Cell 8: 1665-1676.
9 Lindmark DG (1988) Giardia lamblia: localization of hydrolase activities in lysosome-like organelles of trophozoites. Experimental parasitology 65: 141-147.
10 Touz MC, Nores MJ, Slavin I, Carmona C, Conrad JT, et al. (2002) The activity of a developmentally regulated cysteine proteinase is required for cyst wall formation in the primitive eukaryote Giardia lamblia. The Journal of biological chemistry 277: 8474-8481.
11 Ward W, Alvarado L, Rawlings ND, Engel JC, Franklin C, et al. (1997) A primitive enzyme for a primitive cell: the protease required for excystation of Giardia. Cell 89: 437-444.

12 Lanfredi-Rangel A, Attias M, de Carvalho TM, Kattenbach WM, De Souza W (1998) The peripheral vesicles of trophozoites of the primitive protozoan Giardia lamblia may correspond to early and late endosomes and to lysosomes. J Struct Biol 123: 225-235.
13 Touz MC, Lujan HD, Hayes SF, Nash TE (2003) Sorting of encystation-specific cysteine protease to lysosome-like peripheral vacuoles in Giardia lamblia requires a conserved tyrosine-based motif. The Journal of biological chemistry 278: 6420-6426.
14 Bockman DE, Winborn WB (1968) Electron microscopic localization of exogenous ferritin within vacuoles of Giardia muris. J Protozool 15: 26-30.
15 Touz MC, Rivero MR, Miras SL, Bonifacino JS (2012) Lysosomal protein trafficking in Giardia lamblia: common and distinct features. Front Biosci (Elite Ed) 4: 1898-1909.
16 Gaechter V, Schraner E, Wild P, Hehl AB (2008) The single dynamin family protein in the primitive protozoan Giardia lamblia is essential for stage conversion and endocytic transport. Traffic (Copenhagen, Denmark) 9: 57-71.
17 Touz MC, Kulakova L, Nash TE (2004) Adaptor protein complex 1 mediates the transport of lysosomal proteins from a Golgi-like organelle to peripheral vacuoles in the primitive eukaryote Giardia lamblia. Molecular biology of the cell 15: 3053-3060.
18 Rivero MR, Vranych CV, Bisbal M, Maletto BA, Ropolo AS, et al. (2010) Adaptor protein 2 regulates receptor-mediated endocytosis and cyst formation in Giardia lamblia. Biochem J 428: 33-45.
19 Rivero MR, Miras SL, Quiroga R, Ropolo AS, Touz MC (2011) Giardia lamblia low-density lipoprotein receptor-related protein is involved in selective lipoprotein endocytosis and parasite replication. Molecular microbiology 79: 1204-1219.
20 Marti M, Li Y, Schraner EM, Wild P, Kohler P, et al. (2003) The secretory apparatus of an ancient eukaryote: protein sorting to separate export pathways occurs before formation of transient Golgi-like compartments. Molecular biology of the cell 14: 1433-1447.
21 Lujan HD, Mowatt MR, Conrad JT, Bowers BNash TE (1995) Identification of a novel Giardia lamblia cyst wall protein with leucine-rich repeats. Implications for secretory granule formation and protein assembly into the cyst wall. The Journal of biological chemistry 270: 29307-29313.
22 Mowatt MR, Lujan HD, Cotten DB, Bowers B, Yee J, et al. (1995) Developmentally regulated expression of a Giardia lamblia cyst wall protein gene. Molecular microbiology 15: 955-963.
23 Sun CH, McCaffery JM, Reiner DS, Gillin FD (2003) Mining the Giardia lamblia genome for new cyst wall proteins. The Journal of biological chemistry 278: 21701-21708.
24 Gerwig GJ, van Kuik JA, Leeflang BR, Kamerling JP, Vliegenthart JF, et al. (2002) The Giardia intestinalis filamentous cyst wall contains a novel beta(1-3)-N-acetyl-D-galactosamine polymer: a structural and conformational study. Glycobiology 12: 499-505.
25 Jarroll EL, Manning P, Lindmark DG, Coggins JR, Erlandsen SL (1989) Giardia cyst wall-specific carbohydrate: evidence for the presence of galactosamine. Molecular and biochemical parasitology 32: 121-131.
26 Stefanic S, Morf L, Kulangara C, Regos A, Sonda S, et al. (2009) Neogenesis and maturation of transient Golgi-like cisternae in a simple eukaryote. Journal of cell science 122: 2846-2856.
27 Faso C, Konrad C, Schraner EM, Hehl AB (2012) Export of cyst wall material and Golgi organelle neogenesis in Giardia lamblia depend on endoplasmic reticulum exit sites. Cellular microbiology. doi: 10.1111/cmi.12054.

28 Konrad C, Spycher C, Hehl AB (2010) Selective condensation drives partitioning and sequential secretion of cyst wall proteins in differentiating Giardia lamblia. PLoS pathogens 6: e1000835.
29 Marti M, Regos A, Li Y, Schraner EM, Wild P, et al. (2003) An ancestral secretory apparatus in the protozoan parasite Giardia intestinalis. The Journal of biological chemistry 278: 24837-24848.
30 Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, et al. (2007) Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science (New York, NY 317: 1921-1926.
31 Dacks JB, Walker G, Field MC (2008) Implications of the new eukaryotic systematics for parasitologists. Parasitol Int 57: 97-104.
32 Stefanic S, Palm D, Svard SG, Hehl AB (2006) Organelle proteomics reveals cargo maturation mechanisms associated with Golgi-like encystation vesicles in the early-diverged protozoan Giardia lamblia. The Journal of biological chemistry 281: 7595-7604.
33 Jedelsky PL, Dolezal P, Rada P, Pyrih J, Smid O, et al. (2011) The minimal proteome in the reduced mitochondrion of the parasitic protist Giardia intestinalis. PloS one 6: e17285.
34 Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57.
35 Davids BJ, Reiner DS, Birkeland SR, Preheim SP, Cipriano MJ, et al. (2006) A new family of giardial cysteine-rich non-VSP protein genes and a novel cyst protein. PloS one 1: e44.
36 Chiu PW, Huang YC, Pan YJ, Wang CH, Sun CH (2010) A novel family of cyst proteins with epidermal growth factor repeats in Giardia lamblia. PLoS neglected tropical diseases 4: e677.
37 DuBois KN, Abodeely M, Sakanari J, Craik CS, Lee M, et al. (2008) Identification of the major cysteine protease of Giardia and its role in encystation. The Journal of biological chemistry 283: 18024-18031.
38 Castillo-Romero A, Leon-Avila G, Wang CC, Perez Rangel A, Camacho Nuez M, et al. (2010) Rab11 and actin cytoskeleton participate in Giardia lamblia encystation, guiding the specific vesicles to the cyst wall. PLoS neglected tropical diseases 4: e697.
39 Davids BJ, Gilbert MA, Liu Q, Reiner DS, Smith AJ, et al. (2011) An atypical proprotein convertase in Giardia lamblia differentiation. Molecular and biochemical parasitology 175: 169-180.
40 Davids BJ, Mehta K, Fesus L, McCaffery JM, Gillin FD (2004) Dependence of Giardia lamblia encystation on novel transglutaminase activity. Molecular and biochemical parasitology 136: 173-180.
41 Hehl AB, Marti M, Kohler P (2000) Stage-specific expression and targeting of cyst wall protein-green fluorescent protein chimeras in Giardia. Molecular biology of the cell 11: 1789-1800.
42 Hehl AB, Marti M (2004) Secretory protein trafficking in Giardia intestinalis. Molecular microbiology 53: 19-28.
43 Elias EV, Quiroga R, Gottig N, Nakanishi H, Nash TE, et al. (2008) Characterization of SNAREs determines the absence of a typical Golgi apparatus in the ancient eukaryote Giardia lamblia. The Journal of biological chemistry 283: 35996-36010.
44 McCaffery JM, Faubert GM, Gillin FD (1994) Giardia lamblia: traffic of a trophozoite variant surface protein and a major cyst wall epitope during growth, encystation, and antigenic switching. Experimental parasitology 79: 236-249.
45 Svard SG, Meng TC, Hetsko ML, McCaffery JM, Gillin FD (1998) Differentiation-associated surface antigen variation in the ancient eukaryote Giardia lamblia. Molecular microbiology 30: 979-989.
46 Hong W, Lev S (2013) Tethering the assembly of SNARE complexes. Trends Cell Biol. doi: 10.1016/j.tcb.2013.09.006.

47 Morf L, Spycher C, Rehrauer H, Fournier CA, Morrison HG, et al. (2010) The transcriptional response to encystation stimuli in Giardia lamblia is restricted to a small set of genes. Eukaryot Cell 9: 1566-1576.
48 Birkeland SR, Preheim SP, Davids BJ, Cipriano MJ, Palm D, et al. (2010) Transcriptome analyses of the Giardia lamblia life cycle. Molecular and biochemical parasitology 174: 62-65.
49 McArthur AG, Morrison HG, Nixon JE, Passamaneck NQ, Kim U, et al. (2000) The Giardia genome project database. FEMS Microbiol Lett 189: 271-273.
50 Banerjee S, Cui J, Robbins PW, Samuelson J (2008) Use of Giardia, which appears to have a single nucleotide-sugar transporter for UDP-GlcNAc, to identify the UDP-Glc transporter of Entamoeba. Molecular and biochemical parasitology 159: 44-53.
51 Oppermann U, Filling C, Hult M, Shafqat N, Wu X, et al. (2003) Short-chain dehydrogenases/reductases (SDR): the 2002 update. Chemico-biological interactions 143-144: 247-253.
52 Macechko PT, Steimle PA, Lindmark DG, Erlandsen SL, Jarroll EL (1992) Galactosamine-synthesizing enzymes are induced when Giardia encyst. Molecular and biochemical parasitology 56: 301-309.
53 Lanfredi-Rangel A, Attias M, Reiner DS, Gillin FD, De Souza W (2003) Fine structure of the biogenesis of Giardia lamblia encystation secretory vesicles. J Struct Biol 143: 153-163.
54 Faso C, Hehl AB (2011) Membrane trafficking and organelle biogenesis in Giardia lamblia: use it or lose it. International journal for parasitology 41: 471-480.
55 Faso C, Bischof S, Hehl AB (2013) The proteome landscape of Giardia lamblia encystation. PLoS One in press.
56 Vazquez-Martinez R, Diaz-Ruiz A, Almabouada F, Rabanal-Ruiz Y, Gracia-Navarro F, et al. (2012) Revisiting the regulated secretory pathway: from frogs to human. Gen Comp Endocrinol 175: 1-9.
57 Turkewitz AP (2004) Out with a bang! Tetrahymena as a model system to study secretory granule biogenesis. Traffic (Copenhagen, Denmark) 5: 63-68.
58 Reiner DS, McCaffery M, Gillin FD (1990) Sorting of cyst wall proteins to a regulated secretory pathway during differentiation of the primitive eukaryote, Giardia lamblia. Eur J Cell Biol 53: 142-153.
59 Chatterjee A, Carpentieri A, Ratner DM, Bullitt E, Costello CE, et al. (2010) Giardia cyst wall protein 1 is a lectin that binds to curled fibrils of the GalNAc homopolymer. PLoS pathogens 6: e1001059.
60 Abdul-Wahid A, Faubert GM (2004) Similarity in cyst wall protein (CWP) trafficking between encysting Giardia duodenalis trophozoites and CWP-expressing human embryonic kidney-293 cells. Biochemical and biophysical research communications 324: 1069-1080.
61 Beuret N, Stettler H, Renold A, Rutishauser J, Spiess M (2004) Expression of regulated secretory proteins is sufficient to generate granule-like structures in constitutively secreting cells. The Journal of biological chemistry 279: 20242-20249.
62 Kim T, Tao-Cheng JH, Eiden LE, Loh YP (2001) Chromogranin A, an "on/off" switch controlling dense-core secretory granule biogenesis. Cell 106: 499-509.
63 Voorberg J, Fontijn R, Calafat J, Janssen H, van Mourik JA, et al. (1993) Biogenesis of von Willebrand factor-containing organelles in heterologous transfected CV-1 cells. EMBO J 12: 749-758.
64 Scott DC, Schekman R (2008) Role of Sec61p in the ER-associated degradation of short-lived transmembrane proteins. The Journal of cell biology 181: 1095-1105.
65 Lopez AB, Sener K, Trosien J, Jarroll EL, van Keulen H (2007) UDP-N-acetylglucosamine 4'-epimerase from the intestinal protozoan Giardia intestinalis lacks UDP-glucose 4'-epimerase activity. The Journal of eukaryotic microbiology 54: 154-160.

66 Samuelson J, Banerjee S, Magnelli P, Cui J, Kelleher DJ, et al. (2005) The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. Proc Natl Acad Sci U S A 102: 1548-1553.
67 Cao Z, Li C, Higginbotham JN, Franklin JL, Tabb DL, et al. (2008) Use of fluorescence-activated vesicle sorting for isolation of Naked2-associated, basolaterally targeted exocytic vesicles for proteomics analysis. Mol Cell Proteomics 7: 1651-1667.
68 Gauthier DJ, Sobota JA, Ferraro F, Mains RE, Lazure C (2008) Flow cytometry-assisted purification and proteomic analysis of the corticotropes dense-core secretory granules. Proteomics 8: 3848-3861.
69 Brunner Y, Schvartz D, Coute Y, Sanchez JC (2009) Proteomics of regulated secretory organelles. Mass spectrometry reviews 28: 844-867.
70 Lee YH, Tan HT, Chung MC (2010) Subcellular fractionation methods and strategies for proteomics. Proteomics 10: 3935-3956.
71 Jerlstrom-Hultqvist J, Einarsson E, Xu F, Hjort K, Ek B, et al. (2013) Hydrogenosomes in the diplomonad Spironucleus salmonicida. Nat Commun 4: 2493.
72 Aguilar RM, Bustamante JJ, Hernandez PG, Martinez AO, Haro LS (1999) Precipitation of dilute chromatographic samples (ng/ml) containing interfering substances for SDS-PAGE. Analytical biochemistry 267: 344-350.
73 Jimenez-Garcia LF, Zavala G, Chavez-Munguia B, Ramos-Godinez Mdel P, Lopez-Velazquez G, et al. (2008) Identification of nucleoli in the early branching protist Giardia duodenalis. International journal for parasitology 38: 1297-1304.

Supplementary material

Table S1: Protein identification by mass spectrometry

Raw data exported from proteome software (Scaffold version 3.0). Identification probabilities, quantification values, and number of unique peptides (files 1-3) are indicated for all detected proteins. For each of the 1281 proteins identified, the following information is provided: Product description (column B) and GeneID (column C) according to the *G. lamblia* genome database (GiardiaDB), molecular weight (column D), and the T-test scores (column E). For ESV-enriched fractions (E1-E3, columns F-H) and PV-enriched fractions (P1-P3, columns I-K) the protein probabilities (first worksheet), the quantitative values (second worksheet) and the number of unique peptides (third worksheet) are indicated.

This table is too large to be incorporated into this document. Information is available on request and will be available online as soon as the manuscript is published.

Table S2: ESV and PV candidate list with additional information

For each of the 72 ESV and 82 PV candidates, the following information is provided: protein category (column B), GeneID (column C) and product description (column E) according to the *G. lamblia* genome database, NCBI reference number (column D), manual re-annotation (column F) and the prediction tools it is based on (column G), number of transmembrane domains (column H), signal peptide (column I), significant stage-specific up-regulation of transcription (column J), localization of HA-tagged variants determined in this study (column K), and literature information (column L). *TMD:* Transmembrane domains; *SP:* Signal peptide; *ER:* Endoplasmic Reticulum; *ESV:* Encystation specific vesicles; *PV:* Peripheral vesicles; *CYT:* Cytosol *asterisk:* prediction tools return different results.

This table is too large to be incorporated into this document. Information is available on request and will be available online as soon as the manuscript is published.

Table S3: Oligonucleotide primer sequences

Primers used for cloning of expression constructs. Sequences are in 5' to 3' orientation, restriction sites are marked in bold. *pCWP1:* inducible promoter of *G. lamblia* cyst wall protein 1; *pendo:* endogenous promoter, *HA:* hemagglutinin tag.

Construct	**Forward primer (5' to 3' direction)**		**Reverse primer (5' to 3' direction)**	
pCWP1-15483HA	15483_s_NsiI	CG**ATGCAT**AACGAGACTGTCCGTTTCCTTC	15483_as_HA_PacI	CG**TTAATTAA**TCACGCGTAGTCTGGGACATCGT ATGGGTAATCTTGTTTTTTATTCTTACTGCG
pCWP1-10221HA	10221_s_AvrII	GC**CCTAGG**ATGGAAAAATACCACGGGCTG	10221_as_SphI	GA**GCATGC**TTCACTGTGTATGGCCGCGAC
Pendo-10221HA	p10221_s_XbaI	CG**TCTAGA**GAGTATCTGCACCTCCAGTG	10221_as_SphI	GA**GCATGC**TTCACTGTGTATGGCCGCGAC
pCWP1-88581HA	88581_s_AvrII	GC**CCTAGG**ATGAAGATTTCCTTGCTCTGTTC	88581_as_SphI	GA**GCATGC**GAGGAAGGGGAAAATGCAGAAC
Pendo-88581HA	p88581_s_XbaI	CG**TCTAGA**GTTTAACCTAATGACCGTTATTG	88581_as_SphI	GA**GCATGC**GAGGAAGGGGAAAATGCAGAAC
pCWP1-22135HA	22136_s_AvrII	GC**CCTAGG**ATGACTGATACCGAGTCACAG	22136_as_SphI	GA**GCATGC**GAAGCCACTGGCAATATTAATAAAA AG
Pendo-22135HA	p22136_s_XbaI	CG**TCTAGA**CAGAATCCAAGATGCTGGTC	22136_as_SphI	GA**GCATGC**GAAGCCACTGGCAATATTAATAAAA AG
pCWP1-11299HA	11299_s_AvrII	GC**CCTAGG**ATGCAAGTCGAAGACCCTGGTG	11299_as_SphI	GA**GCATGC**TTTTTTATTGTACGGACAATTATCAA C
Pendo-11299HA	p11299_s_XbaI	CG**TCTAGA**CTGTAGTGGAGTTGTCGGTC	11299_as_SphI	GA**GCATGC**TTTTTTATTGTACGGACAATTATCAA C
pCWP1-11595HA	11595_BclI_s	CA**TGATCA**ATGACGACCGCGGCGTC	11595_SphI_as	GA**GCATGC**ATCCTTGCGTGTTCTAGCTC
Pendo-11595HA	p11595_XbaI_s	CG**TCTAGA**CTTGCCCAAGCAGCGGAC	11595_SphI_as	GA**GCATGC**ATCCTTGCGTGTTCTAGCTC
pCWP1-32419HA	32419_AvrII_s	GC**CCTAGG**ATGAAGCACATAGGTCTCCAC	32419_SphI_as	GA**GCATGC**GCGAAGAGAAAGGGGCTGC
pendo-32419HA	P32419_Xba_for	GC**TCTAG**AGCGCATGCTAGGAATCGC	32419_SphI_as	GA**GCATGC**GCGAAGAGAAAGGGGCTGC
pCWP1-15956HA	15956_AvrII_s	GC**CCTAGG**ATGGTTCTTCAACTCCGCAG	15956_PacI_HA_as	CG**TTAATTAA**TCACGCGTAGTCTGGGACATCGT ATGGGTACATTGACCACACCATGACAC
Pendo-15956HA	p15956_XbaI_s	CG**TCTAGA**CAGTTTGCCCAACACGCTTC	15956_PacI_HA_as	CG**TTAATTAA**TCACGCGTAGTCTGGGACATCGT ATGGGTACATTGACCACACCATGACAC
pCWP1-7207HA	7207_AvrII_s	GC**CCTAGG**ATGCAACAGCAGCAAGACCCAC	7207_SphI_as	GA**GCATGC**TAAACGATATTGGAGAACACTCTC
Pendo-7207HA	7207_SpeI_s	GC**ACTAGT**CAGATTCTGTGCTATTTGAATAAG	7207_SphI_as	GA**GCATGC**TAAACGATATTGGAGAACACTCTC
pCWP1-7350	7350_AvrII_s	GC**CCTAGG**ATGGTGAAAGTGGCCGAAG	7350_HA_PacI_as	CGTTAATTAATCACGCGTAGTCTGGGACATCGT ATGGGTAGGAAACACACAAAGCAGGAATTC
Pendo-7350	7350_SpeI_s	GC**ACTAGT**CTCACGCGAACTGTAGCCTG	7350_HA_PacI_as	CGTTAATTAATCACGCGTAGTCTGGGACATCGT ATGGGTAGGAAACACACAAAGCAGGAATTC
pCWP1-14458HA	14458_AvrII_s	GC**CCTAGG**ATGAGCAAACATCTGGCCATTG	14458_SphI_as	GA**GCATGC**CTTGCAAAGGATCTTCCTCATAG
Pendo-14458HA	14458_XbaI_s	CG**TCTAGA**CTTTGGCGGCTTTATGGCAG	14458_SphI_as	GA**GCATGC**CTTGCAAAGGATCTTCCTCATAG

pCWP1-25205HA	25205_AvrII_s	GC**CCTAGG**ATGGCCGGAAATTACCCTGTC	25205_HA_PacI_as	CGTTAATTAATCACGCGTAGTCTGGGACATCGT ATGGGTAGATTTCCATTAGGCCATAATG
Pendo-25205HA	25205_SpeI_s	GC**ACTAGT**GTACGTTCACACTCTTTACCAG	25205_HA_PacI_as	CGTTAATTAATCACGCGTAGTCTGGGACATCGT ATGGGTAGATTTCCATTAGGCCATAATG
Pendo-15156HA	p15156_XbaI_s	CG**TCTAGA**GTACGCTTTTGGCGAGTCAGAG	15156_SphI_as	GA**GCATGC**CTTTGACATCTTTGCCTTCTGAATC
Pendo-10780HA	p10780_XbaI_s	CG**TCTAGA**GTTTCCGGAATAATGTTTGTCAG	10780_SphI_as	GA**GCATGC**CTTCTGTCTGTACCTCTCTAC
pCWP1-9157HA	9157_SpeI_s	GC**ACTAGT**ATGAACAGCGGGGCTGACTCTG	9157_SphI_as	GA**GCATGC**TGCGGATACTTCGTAGTTCCGTAG
Pendo-9157HA	p9157_XbaI_s	CG**TCTAGA**GATCAGCGCGTCTAATGGCGTG	9157_SphI_as	GA**GCATGC**TGCGGATACTTCGTAGTTCCGTAG
pCWP1-87926HA	87926_AvrII_s	GC**CCTAGG**ATGTTTGCGGCAGACTCTCAAAAC	87926_SphI_as	GA**GCATGC**CTCGCCCTTCTCCCTCAAG
Pendo-87926HA	p87926_AvrII_s	GC**CCTAGG**GTCGTGAATGACGATGAGACTG	87926_SphI_as	GA**GCATGC**CTCGCCCTTCTCCCTCAAG
Pendo-8382HA	8382_SpeI_s	GC**ACTAGT**ATGCTGTCTTTCTGGGTTTATC	8382_HA_PacI_as	CG**TTAATTAA**TCACGCGTAGTCTGGGACATCGT ATGGGTACTTAAGAGCACACTTTATTAGGAAC
Pendo-7982HA	7982_SpeI_s	GC**ACTAGT**ATGCAGAACCACTCGTTCCTC	7982_HA_PacI_as	CG**TTAATTAA**TCACGCGTAGTCTGGGACATCGT ATGGGTATAGACCAATCTCCTTGACATC
pCWP1-96994HA	96994_AvrII_s	AC**CCTAGG**ATGACCGATTTTG ATGCACC	96994_Sph1_as	GT**GCATGC**CTTTTTCACGGCACGGATG
pCWP1-15104HA	15104_AvrII_s	CA**CCTAGG**GCCATTGAAACTTGAGGGAG	15104_Sph1_as	GT**GCATGC**TGCTTTGAGGAACACCCCC
pCWP1-16521HA	16521_AvrII_s	CG**CCTAGG**ATGAGTTATGCAAAGCAGGCG	16521_Sph1_as	CT**GCATGC**CAACAACTCGCCGATAGCG
pCWP1-15472HA	15472_AvrII_s	GC**CCTAGG**ATGGGGAAGATAGATCTTACC	15472_Sph1_as	CG**GCATGC**CTCGACGCTCTGGGCCA
pCWP1-15339HA	15339_AvrII_s	CG**CCTAGG**ATGGGCTCAACAACAAACATT	15339_HA_PacI_as	GC**TTAATTAA**TCACGCGTAGTCTGGGACATCGT ATGGGTAATTTCTGTTCAGGCCAGTAAC
Pendo-4270HA	4270_Xba1_s	GC**TCTAGA**TGGCACAGAGCGC GTCAAC	4270_HA_PacI_as	GC**TTAATTAA**CTACGCGTAGTCTGGGACATCGT ATGGGTAGATCGGTAAATCCACAATATATGC
pCWP1-8559HA	8559_NsiI_s	CG**ATGCAT**ATGTCGACAACATCCAGGAATC	8559_HA_PacI_as	CG**TTAATTAA**TCACGCGTAGTCTGGGACATCGT ATGGGTACTGCTGCACGCAAATGTTCC
pCWP1-40224HA	40224_AvrII_s	GC**CCTAGG**ATGAAGTCTGCGTTCAACCC	40224_Sph1_as	GC**GCATGC**TAAGAACTTAGGTTGATACAGC

Figure S1: Workflow

CWP3-GFP expressing cells at 13 hours p.i. **(A)** and wild type trophozoites after endocytic uptake of the fluid phase dye Dextran-AlexaFluor-647 **(B)** were disrupted by sonication and passed through a 5µm filter. The cleared microsome fractions were mixed **(C)** and organelles were simultaneously enriched by flow cytometry-assisted organelle sorting (FAOS) **(D)**. Sample preparation and organelle sorting were performed in biological triplicates. Protein precipitates of organelle-enriched fractions were separated by 1D-SDS-PAGE and analyzed by mass spectrometry (MS) **(E)**, resulting in 3 ESV and PV mass spectrometry datasets, each **(F)**. Contaminating proteins were identified by intersecting the ESV and PV MS-datasets **(G)**. A detailed description of the intersection can be found in figure S3. In silico data filtration, i.e., removal of the data intersection **(H)** revealed ESV-organelle **(J)** and PV-organelle **(K)** specific datasets.

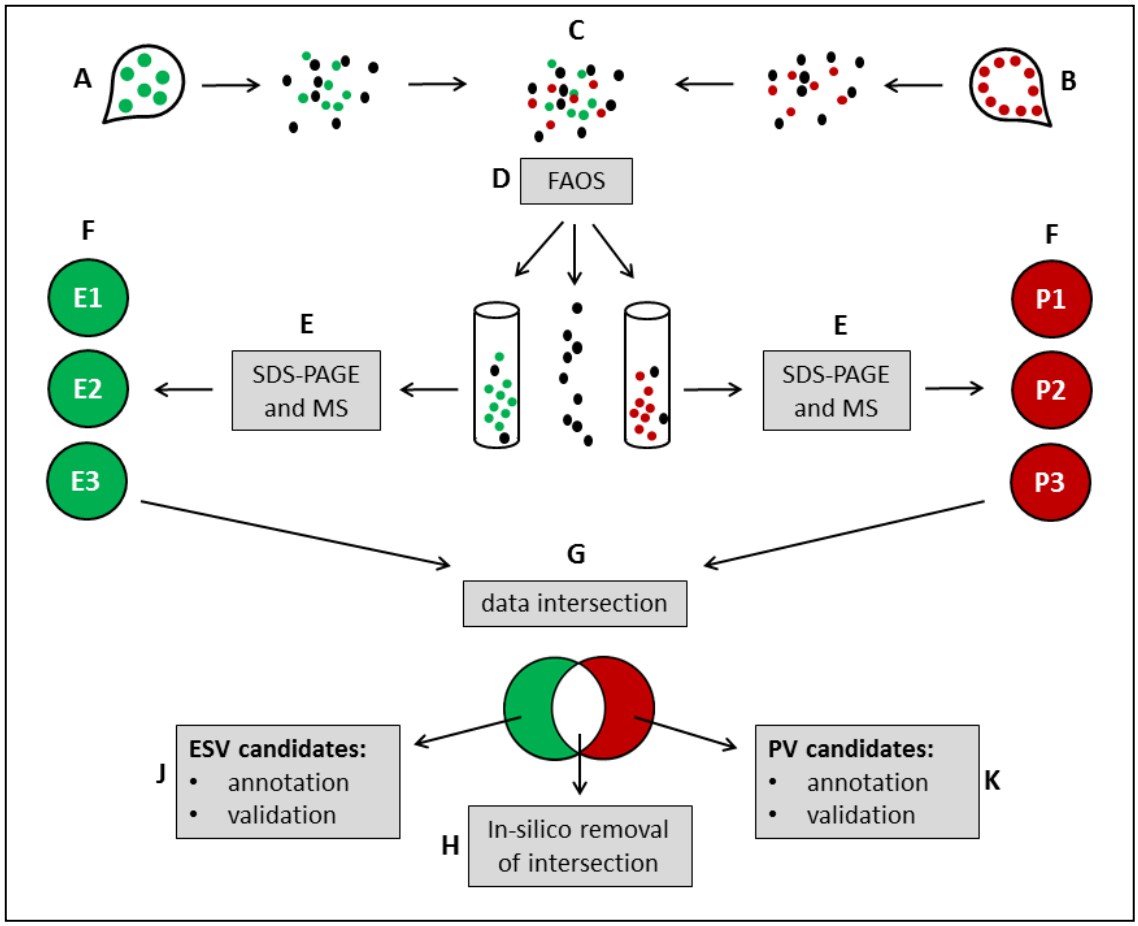
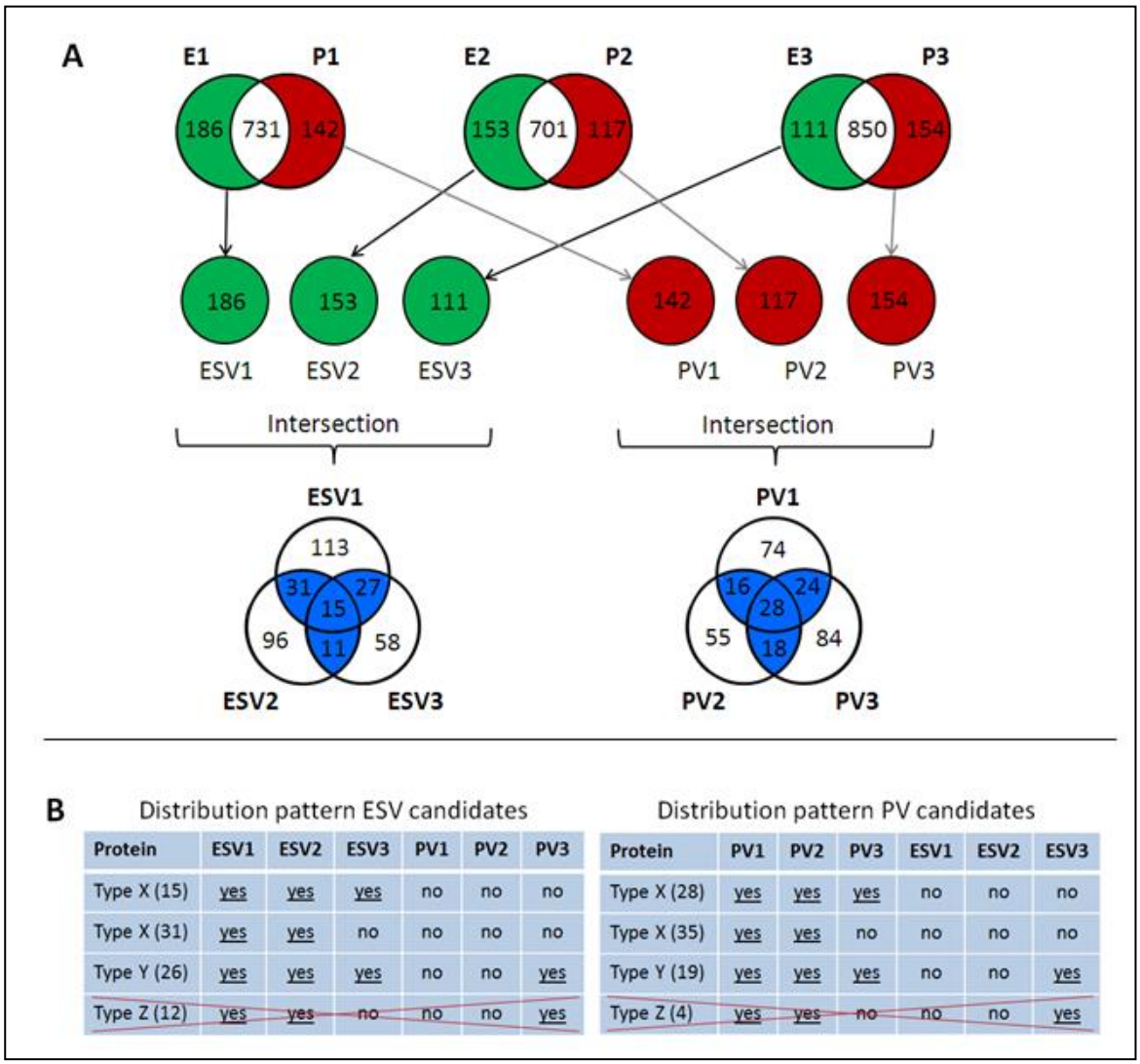


Figure S2: Generation of the MS data intersection

Mass spectrometry datasets of ESV-enriched (E1, E2, E3) and PV-enriched (P1, P2, P3) fractions of each replicate were intersected separately (top). The numbers stand for the proteins detected by mass spectrometry. Removal of the intersection revealed proteins exclusively detected in ESV-enriched fractions (middle, left) or PV-enriched fractions (middle, right). From these lists, only proteins occurring in at least two lists were accepted (bottom, blue). The proteins were further analyzed according to their distribution pattern in the six organelle-enriched fractions (B). **B)** Schematic representation of the protein distribution pattern in ESV- and PV-enriched fractions. ESV candidates (left): proteins of type X were detected exclusively and in at least two of three ESV fractions, proteins of type Y were detected in all ESV fractions and in one PV fraction, proteins of type Z were detected in only two ESV fractions and in one PV fraction. The same is true vice-versa for PV candidates (right). Type Z proteins were removed, resulting in 72 ESV and 82 PV candidate proteins. The respective protein numbers are indicated in brackets.



B

Distribution pattern ESV candidates

Protein	ESV1	ESV2	ESV3	PV1	PV2	PV3
Type X (15)	yes	yes	yes	no	no	no
Type X (31)	yes	yes	no	no	no	no
Type Y (26)	yes	yes	yes	no	no	yes
Type Z (12)	yes	yes	no	no	no	yes

Distribution pattern PV candidates

Protein	PV1	PV2	PV3	ESV1	ESV2	ESV3
Type X (28)	yes	yes	yes	no	no	no
Type X (35)	yes	yes	no	no	no	no
Type Y (19)	yes	yes	yes	no	no	yes
Type Z (4)	yes	yes	no	no	no	yes

Figure S3: Conserved short chain dehydrogenases (SDH) motifs in Gl8382, Gl7982 and human GALE

Protein sequences of *G. lamblia* Gl7982 (cytoplasmic GALE, [52], *G. lamblia* Gl8382 (putative ER-GALE), and the human GALE (hGALE) were analyzed manually. All conserved sequences required for hGALE function [51] are present in both Giardia GALEs and listed in the table. A conserved PG motif, which is required for the direction of the reaction, is only present in the Giardia ER-GALE. An N-terminal integral membrane domain in the ER-GALE shifts the conserved motif positions for about 40 amino acids towards the C-terminus, compared to the cytoplasmic GALE and hGALE.

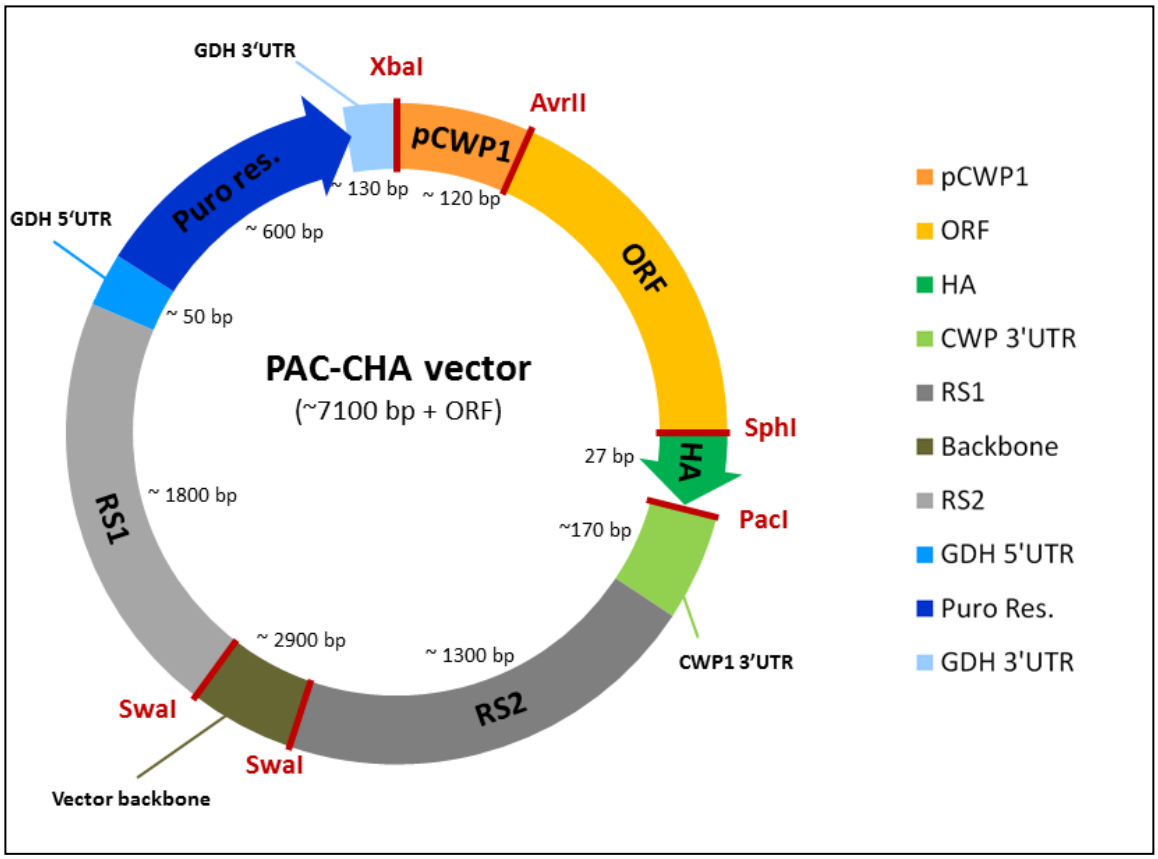
```
Gl8382   MLSFWVYLLLLFTTVLLARISRLASERKSARRSLLTGDSGLCVITGGGSGLGEALAHEYAALGHPLLLIGRKHENLERAKHALIEKFGEKASVTVLAADL
Gl7982   ------------------------------MQNHSFLGKTVLITGGCGFIGSHFVEACHVLGMTVYVLDNLSSGKNVFKTTSDCSSSLVYTIGDIRDKA
hGALE    -----------------------------------MAEKVLVTGGAGYIGSHTVLELLEAGYLPVVIDNFHNAFRGGGSLPESLRRVQELTGRSVEFE

Gl8382   EVESECIKVVEYLRDKSVFHLVNNAGVGQVSADISVSDARSIASINLTAPAVLSAGIHAGRYIFISSPQGLLPMPNRCVYAASKAGIHNLAESMRLDGMN
Gl7982   IFSRLPQKIDFVIHLAAAVSVAESVTNPQKYMLTNVEGSRNVFQYAVDAKASAVLSASTAAYYGDCGKSAITEAFPYGGISPYAESKMEMERLGAEFQKT
hGALE    EMDILDQGALQRLFKKYSFMAVIHFAGLKAVGESVQKPLDYYRVNLTGTIQLLEIMKAHGVKNLVFSSSATVYGNPQYLPLDEAHPTGGCTNPYGKSKFF

Gl8382   VTTAYPGWIAILRLRMNARGNGLPNPQARTARSAEDVAKEIVSAALSGKRACCLDAKTRIVAILYSVWPEAAEFLIKCALK--------------------
Gl7982   SRCRFIFCRFFNVYGPRQDPSSPYTGVMSIFMDRCAARKPITIFGTGEQTRDFVFIKDLIVAAINLLGQLDKFPIGADAVQQNDPEEVQRSAYTGEGVYP
hGALE    IEEMIRDLCQADKTWNAVLLRYFNPTGAHASGCIGEDPQGIPNNLMPYVSQVAIGRREALNVFGNDYDTEDGTGVRDYIHVVDLAKGHIAALRKLKEQCG

Gl8382   ----------------------------------------------------------------------------------------------------
Gl7982   TVFNIGSGISISVNELAELAKIVSGRHEVEIVHGEPRSGDILHSLSDCTRIRNATGWSASTTLRVGMSETWGWAAGEISYLSGDLVRVLENELKIDGVPV
hGALE    CRIYNLGTGTGYSVLQMVQAMEKASGKKIPYKVVARREGDVAACYANPSLAQEELGWTAALGLDRMCEDLWRWQKQNPSGFGTQA

Gl8382   ----------------
Gl7982   AKSLCGKDADVKEIGL
hGALE    ----------------
```

Conserved motif	Position	Function	Putative ER-GALE (Gl8382)	Cytosolic GALE (Gl7982)	Human GALE (hGALE)
TGxxxGxG	12 to 19	Co-enzyme binding, maintenance of central β-sheet	TGGGSGLG Position 45 - 52	TGGCGFIG Position 14 - 21	TGGAGYIG Position 8 - 15
D frequently found	60	Stabilization of adenine ring pocket, weak binding to coenzyme	D Position 99	D Position 64	D Position 66
NNAG (or Ala88) frequently found	86-89	Stabilization of central β-sheet	NNAG Position 123 - 126	A Position 79, probably	A Position 89, probably

Conserved motif	Position	Function	Putative ER-GALE (Gl8382)	Cytosolic GALE (Gl7982)	Human GALE (hGALE)
N frequently found	111	Active site	NNAG Position 124, probably	N Position 104, probably	N Position 108, probably
S-Y-K	138, 151, 155	Active site	SAGIHAGRYIFISSPQ GLLPMPNRCVYAASK Position 154, 180, 184	STAAYYGDCGKSAIT EAFPYGGISPYAESK Position 127, 152, 156	SATVYGNPQYLPLDE AHPTGGCTNPYGKSK Position 132, 157, 161
N	179	Connection of substrate binding loop and active site	N Position 200	N Position 181	N Position 179
PG least conserved	183-184	Reaction direction, presumably involved in co-factor interactions	PG Position 207	Not present	Not present
T least conserved	188	H-bonding to carboxamide of nicotinamide ring	T Position 211	T Position 193	T Position 189
C	307		C Position 277	C Position 217, probably	C Position 307

Figure S4: Vector map

Schematic depiction of the vector used for candidate cloning. *pCW1:* putative promoter region of cyst wall protein 1 (GL50803_5638); *ORF:* open reading frame; HA: hemagglutinin tag; *CWP 3'UTR:* 3' untranslated region of cyst wall protein 1 (GL50803_5638); *RS1/2:* recombination sites 1 (GL50803_17200) and 2 (GL50803_93938); *GDH 5'/3' UTR:* 5' and 3' untranslated regions of glutamate dehydrogenase (GL50803_21942); *Puro Res.:* puromycin N-acetyltransferase.



Text S1: Detailed description of mass spectrometry analysis

Detailed description of SDS-PAGE, sample preparation, mass spectrometry analysis, database search and protein identification.

SDS-PAGE and sample preparation for mass spectrometry

After thawing, protein samples was boiled for 5 minutes and centrifuged for 1 minute at 16'100xg at room temperature to pellet undissolved material. Samples were separated by 1D-SDS-PAGE using precast 12% Tris-Glycine gels (Invitrogen, IM6000). 4 a- 8 ul of supernatant were loaded, depending on the estimated amount of protein in the respective sample determined in a preceding test run. The gel was incubated in fixing solution (20% (v/v) methanol, 1% (v/v) 85% phosphoric acid, 79% water) for 15 minutes and stained in Roti Blue (20% (v/v) methanol, 20% (v/v) Roti Blue concentrate, 60% water) over

night. After destaining the gel for 3 x 5 minutes in washing solution (25% (v/v) methanol, 75% water), each gel band was cut into 21 slices. Reduction of disulfide bridges was performed by covering the vacuum dried gel pieces with sufficient 10mM DTT in 25mM ammonium bicarbonate, pH 8, and incubation for 45 minutes at 60°C. For alkylation of cysteines DTT was removed, and 50mM iodacetamide in 25mM ammonium bicarbonate, pH8, was added for one hour at room temperature. Gel pieces were washed twice in 50% acetonitrile and airdried. In-gel digestion was performed in 25mM ammonium bicarbonate pH 8 containing 50 ng trypsin (Roche, 03708985001). After 5 minutes, gel pieces were overlaid with 25mM ammonium bicarbonate and incubated over night at 37°C. Peptides were extracted by addition of 50% acetonitrile / 5% trifluoroacetic acid to the gel pieces and subsequent removal of supernatant. The step was repeated twice. Peptides were dried and desalted using C18 ZipTips (Millipore, ZTC18S960) and the following solutions: wetting solution (100% Ethanol), washing solution (3% acetonitrile; 0.1% trifluoroacetic acid) and elution solution (80% acetonitrile; 0.1% trifluoroacetic acid). After drying, peptides were resuspended in 3% acetonitrile and 0.2% formic acid and incubated for 15 minutes at room temperature.

Mass spectrometry

Mass Spectrometry analysis was performed at the Functional Genomics Center Zurich. Samples were analyzed on a LTQ-Orbitrap XL mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to an Eksigent-Nano-HPLC system (Eksigent Technologies, Dublin, CA, USA). Solvent composition at the two channels was 0.2% formic acid, 1% acetonitrile for channel A and 0.2% formic acid, 80% acetonitrile for channel B. Peptides were resuspended in 3% acetonitrile and 0.2% formic acid and loaded on a self-made tip column (75 µm × 70 mm) packed with reverse phase C18 material (AQ, 3 µm 200 Å, Bischoff GmbH, Leonberg, Germany) and eluted with a flow rate of 200 nl per min by a gradient from 3 to 10% of B in 5 min, 48% B in 55 min, 97% B in 60 min. Full-scan MS spectra (300−2000 m/z) were acquired with a resolution of 60 000 at 400 m/z after accumulation to a target value of 500 000.
Collision induced dissociation (CID) MS/MS spectra were recorded in data dependent manner in the ion trap from the six most intense signals above a threshold of 500, using a normalized collision energy of 35% and an activation time of 30 ms. Charge state screening was enabled and singly charge states were rejected. Precursor masses selected twice for MS/MS were excluded for further selection for 120s. The exclusion window was set to 20 ppm, while the size of the exclusion list was set to a maximum of 500 entries. Samples were acquired using internal lock mass calibration set on m/z 429.088735 and 445.120025.

Database search and protein identification

The raw-files from the mass spectrometer were converted into Mascot generic files (mgf) with Mascot Distiller software 2.4.2.0 (Matrix Science Ltd.,London, UK). The peak lists were searched using Mascot Server 2.3 against the *Giardia lamblia* database (http://tinyurl.com/37z5zqp) with a concatenated decoy database supplemented with contaminants and The Arabidopsis Information Resource (TAIR9) protein database and the Swissprot database (to increase the database's size). The final database included 79141

entries. The parameters for precursor tolerance and fragment ion tolerance were set to ± 5 ppm and ± 0.8 Da, respectively. The identification results were loaded into Scaffold 3.0 (Proteome Software, Portland, US) and filtered for a minimal mascot score of 20 for peptide probability, a protein probability greater than 80% and a minimum of 2 unique peptides per protein.

Text S2A: DAVID functional annotation clustering of data intersection

Functional annotation clustering of the data intersection using the DAVID bioinformatics tool (http://david.abcc.ncifcrf.gov/home.jsp). For each cluster, the enrichment score is given at the top, and functional groups (left) and protein counts (right) within the respective cluster are listed. The 1059 putative contaminants of the data intersection are categorized into 56 annotation clusters.

This table is too large to be incorporated into this document. Information is available on request and will be available online as soon as the manuscript is published.

Text S2B: DAVID functional annotation clustering of the *G. lamblia* genome

Functional annotation clustering of the *G. lamblia* genome using the DAVID bioinformatics tool (http://david.abcc.ncifcrf.gov/home.jsp). For each cluster, the enrichment score is given at the top, and functional groups (left) and protein counts (right) within the respective cluster are listed. Clustering of all predicted proteins in the *G. lamblia* genome (5150 validated genes) using 10 random gene lists containing 3'000 genes each.

This table is too large to be incorporated into this document. Information is available on request and will be available online as soon as the manuscript is published.

2. The Cre/loxP system in *Giardia lamblia*: genomic manipulations in a tetraploid protozoan

This study represents a smaller part of my PhD work which I started in the third year. The project idea, designs and experimental strategies were developed by me and my direct supervisor Prof. Adrian B. Hehl, who supported me in all steps during this project.

All experimental work was performed by me, except for some of the final analysis steps which were performed by Dr. Carmen Faso (figure 4C and figure 5 middle and bottom panels). Manuscript draft writing and figure assembly was performed by me with the support of Dr. Carmen Faso. The manuscript is currently prepared for submission.

The combined results of this study and the main study of my PhD (Proteomics of secretory and endocytic organelles in *Giardia lamblia*) gave rise to two follow-up projects performed by Dr. Carmen Faso and Jacqueline Ebneter in our laboratory.

The Cre/loxP System in *Giardia lamblia*: Genetic Manipulations in a Tetraploid Protozoan

Petra B. Wampfler, Carmen Faso and Adrian B. Hehl*

Institute of Parasitology, University of Zurich (ZH), Switzerland

Key words: Cre/loxP, Giardia, tetraploid, genetic manipulation, selectable marker

*To whom correspondence and requests for materials should be addressed
Adrian B. Hehl
Laboratory of Molecular Parasitology
Institute of Parasitology
University of Zurich
Winterthurerstrasse 266a
8057 Zürich, Switzerland
adrian.hehl@access.uzh.ch
Tel. +41 44 635 8526
Fax +41 44 635 8907

Abstract

The bacteriophage P1-derived Cre/loxP system is a fascinating and valuable tool that has revolutionized genetic and cell biological research by making diverse genetic manipulations possible in a wide variety of organisms. The Cre recombinase efficiently mediates recombination between specific 34 bp-sized loxP sites, resulting in excision, insertion, inversion or translocation of loxP-flanked (floxed) DNA sequences. We implemented the Cre/loxP system in *Giardia lamblia* - an evolutionary diverged protozoan whose tetraploid genome organization severely limits the application of reverse genetic approaches to study molecular and cellular processes. We demonstrate that conditional expression of functional Cre recombinase can be used to recycle selectable markers in order to genetically fix insertions/deletions in *G. lamblia* chromosomes. Providing the means for more complex and versatile genetic manipulations, the Cre/loxP technique has the potential to significantly impact the investigation of gene and protein function in *G. lamblia*.

Introduction

Giardia lamblia is a fascinating and unique protozoan model to investigate minimal cellular mechanisms and explore eukaryotic diversity. This intestinal parasite is an early-diverged eukaryote that has undergone substantial reductive evolution to an even simpler organization than the predicted last common eukaryotic ancestor (*1*). Giardia has simplified molecular machineries for most cellular processes and metabolic pathways, and canonical eukaryotic organelles such as mitochondria, peroxisomes, and the Golgi have been minimized or even lost (*2-4*).

The *Giardia* life cycle comprises a motile trophozoite stage which multiplies by binary fission in the gut of animal and human hosts and an environmentally resistant and infections cyst stage which is shed with the feces. As a member of the Diplomonad phylum, *Giardia* trophozoites have two diploid nuclei which are both transcriptionally active, making the cell effectively tetraploid (*5, 6*). While the binuclear state raises interesting questions about the emergence and functionality of each nucleus, Giardia's tetraploid status restricts the application of genetic and cell biological approaches to investigate this intestinal parasite. This situation is further complicated by several additional factors:
i) targeted insertion of linear DNA molecules by double cross-over is efficient in the WB isolate of *G. lamblia* which is the standard laboratory strain (*7*). However, chromosomal insertion of linear DNA occurs only in one of the two nuclei after transfection although both nuclei are able to harbor inserts (*8*);
ii) the repertoire of suitable drug-resistance markers (*7, 9-11*) for the selection of transgenic parasite cells is limited;
iii) most cells in culture are in G2 of the cell cycle, and it is unclear how this affects the efficiency of recombination events within a target nucleus. Accordingly, multiple transfections of transgenic Giardia lines are not routinely feasible. Nevertheless, co-transfected circular plasmids which contain multiple genes of interest and a single resistance marker can be selected as non-integrated episomes leading to co-expression of up to three heterologous proteins in the majority of transgenic parasites (*12*). However, the study of gene function remains limited: complex genetic manipulations requiring multiple chromosomal alterations such as gene knockouts have not been reported so far.

Given the challenge of manipulating a tetraploid genome, current strategies to deplete protein levels for functional analyses in *G. lamblia* primarily target mRNA production, stability, and accessibility to the translation machinery. The recently uncovered RNA interference machinery (*13, 14*) provides a useful but limited tool to investigate gene or protein function. Gene silencing by expression of antisense sequences (*15, 16*) or long-dsRNA using expression vectors (*17, 18*) has been reported. However, the outcome appears to depend largely on a series of factors such as the choice of expression vector, target RNA and length of the antisense-construct, thus resulting in highly variable knock down efficiencies (*16, 18*). Reduction of target protein levels using hammerhead ribozymes or translation blocking morpholinos likewise result in considerable variation of downregulation of expression (*19-21*) or are not feasible at all because of the unusual mode of translation initiation in Giardia (*22*).

In this report, we addressed the challenge of manipulating Giardia's tetraploid genome to facilitate gene function analyses and cell biology studies in this fascinating protozoan by implementing the well-

characterized Cre/loxP system (*23*). Cre (causes recombination) is a 343 amino acid bacteriophage P1-encoded enzyme that mediates DNA recombination at specific target sites termed loxP (locus of crossing over (x) in P1) (*23*). LoxP sites consist of 34 bp DNA sequences containing an 8 bp asymmetrical core flanked by two 13 bp inverted repeats (*24*). Depending on the relative orientation of two identical loxP sites with respect to one another, Cre mediates excision, insertion, inversion, or translocation of loxP-flanked ("floxed") DNA sequences (*25*). Within the last 30 years the Cre/loxP technology has been harnessed as a powerful research tool to tightly regulate the location and timing of genetic manipulations and it has been successfully applied to a wide variety of organisms including other protozoan parasites such as *Toxoplasma gondii*, *Trypanosoma brucei* and *Plasmodium falciparum* (*26-28*).

To test the feasibility of applying the Cre/loxP system to Giardia, we first determined if Cre is functionally expressed in this protozoan by testing Cre-dependent reconstitution of a green fluorescent protein variant (GFP)-encoding open reading frame (ORF) (*29*) interrupted by an mCherry-encoding expression cassette flanked by loxP sites (floxed). We could show that, following nuclear-targeted Cre-recombinase expression, the floxed mCherry cassette is correctly excised in transgenic parasites, resulting in GFP fluorescence. We then demonstrated that the Cre/loxP system can be successfully implemented for the removal of a floxed puromycin drug-resistance cassette in a chromosomally inserted expression vector, thus allowing for the "recycling" of the resistance marker in subsequent rounds of transfection of transgenic parent cell line.

Taken together, the data presented in this report provide proof-of-concept for a considerably broader range of genomic manipulations in *G. lamblia* using the Cre/loxP system and demonstrate that gene knock outs are technically feasible in this tetraploid protozoan.

Results

Conditionally expressed Cre recombinase excises a floxed mCherry expression cassette and reconstitutes a split GFP ORF in *G. lamblia*

To determine if Cre is functionally expressed in Giardia, we developed a reporter system based on Cre-dependent reconstitution of a green fluorescent protein (GFP)-encoding open reading frame (ORF) interrupted by a floxed mCherry-encoding expression cassette (GFPN-mCherryfl-GFPC, figure 1A). First, we generated a *G. lamblia* cell line stably transfected with the GFPN-mcherryfl-GFPC construct (figure 1A). Stable integration of the linearized construct into the intergenic region flanking the triose phosphate isomerase (TPI) gene (*30*) by double cross-over was confirmed by PCR and sequencing (not shown). We performed flow cytometry-based single cell sorting and collected red fluorescent transgenic cells in 96 well plates. Single cell-derived colonies were established in ~30% of all wells, indicating that roughly every third *G. lamblia* cell survived the procedure. Analysis of the single cell-derived colonies by live cell imaging (figure 2A) revealed that mCherry-expressing cells were viable. Flow cytometry demonstrated a 2.4-fold higher relative average fluorescence intensity and clear distinction of the transgenic parasite population compared with wild type control cells (figure 2B).

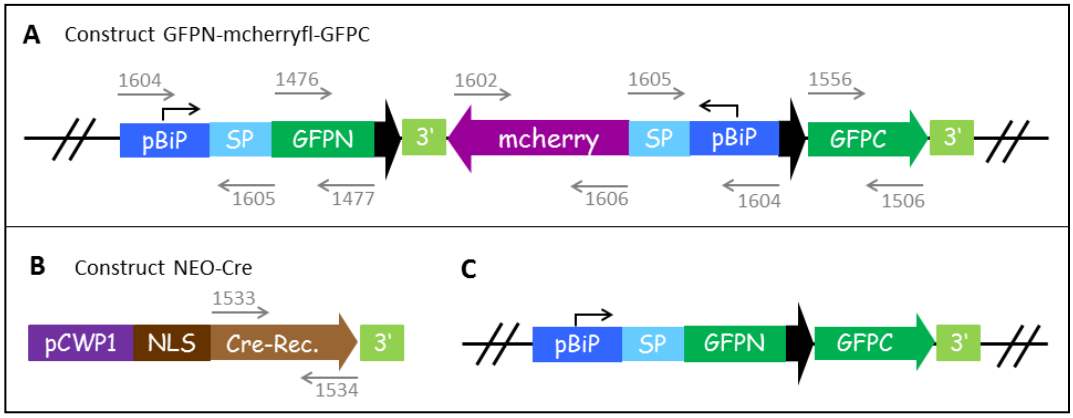


Figure 1: *Schematic illustration of the Cre/loxP reporter system.* **(A)** Cells are stably transfected with construct GFPN-mcherryfl-GFPC. In the absence of Cre recombinase, the construct encodes for a constitutively-expressed full-length ER-targeted mCherry. Black lines represent the site of genomic integration for construct GFPN-mcherryfl-GFPC. *pBiP:* putative promoter sequence of GlHsp70/BiP; *SP:* secretory signal peptide coding sequence from GlHsp70/BiP; *GFPN, GFPC:* N- and C-terminal moieties of the GFP ORF; *3':* 3'UTR of GlCWP1; black arrow: loxP site; *grey arrows:* Oligonucleotides used for construct cloning (table S1). **(B)** The Cre-recombinase is encoded on a separate expression vector NEO-Cre. *CWP1:* cyst wall protein 1 (Gl5638); *3' (green):* 3'UTR of CWP1; *pCWP1:* putative CWP1 promoter region, *NLS:* nuclear localization signal; *Cre-rec:* Cre-recombinase; *grey arrows:* Oligonucleotides used for construct cloning (table S1). **(C)** Transient transfection of GFPN-mcherryfl-GFPC expressing cells with NEO-Cre followed by induction of Cre expression results in loss of the floxed mcherry expression cassette and reconstitution of a full length, constitutively-expressed, ER-targeted GFP ORF. Following excision, a single loxP site remains. Black lines represent the site of genomic integration for construct GFPN-mcherryfl-GFPC.

We introduced Cre into a cell line expressing a mCherry reporter by transfection of a circular, and therefore episomally-maintained, plasmid vector for conditional expression of nuclear-targeted Cre recombinase (construct NEO-Cre, figure 1B). Following G418-based selection of transgenic cells carrying the NEO-Cre- episome, we induced transient Cre expression by activating the cyst wall protein promoter for 1.5 h in encystation medium and placing the cells in standard culture medium again. After propagation for 16 h to allow Cre-mediated reconstruction of the interrupted GFP ORF (figure 1C) we subjected the population to single cell sorting to generate clonal lines for analysis. PCR-based analysis of genomic DNA (gDNA) extracted from clonal lines demonstrated successful reconstruction of the GFP ORF in 6 of 10 selected recombinant cell lines (figure 2C). In addition, Cre-mediated functional reconstitution of the GFP ORF was verified for three single cell-derived colonies by immunofluorescence microscopy (figure 2D) and Western blot analysis (figure 2E) detecting GFP. Taken together, our data demonstrate that Cre recombinase expressed in *G. lamblia* is functional and can be used for specific excision of engineered floxed DNA fragments from chromosomes.

Further PCR and DNA sequencing analyses of the GFP-positive, single cell-derived lines revealed that, although the GFP ORF had been reconstituted, the floxed mCherry cassette was still present (not shown). A possible explanation for this is that the floxed mCherry cassette was maintained in these cells in both integrated and circular form due to the presence of the Cre recombinase. This would have led to the establishment of an equilibrium between insertion and excision of the cassette (*25, 31*), maintained until

the episomal vector containing the *cre* gene is eventually lost in the absence of a positive selection. Given the stringent gating applied to the single cell sorting procedure the presence of cell lines originating from more than one cell was highly unlikely (figure S1).

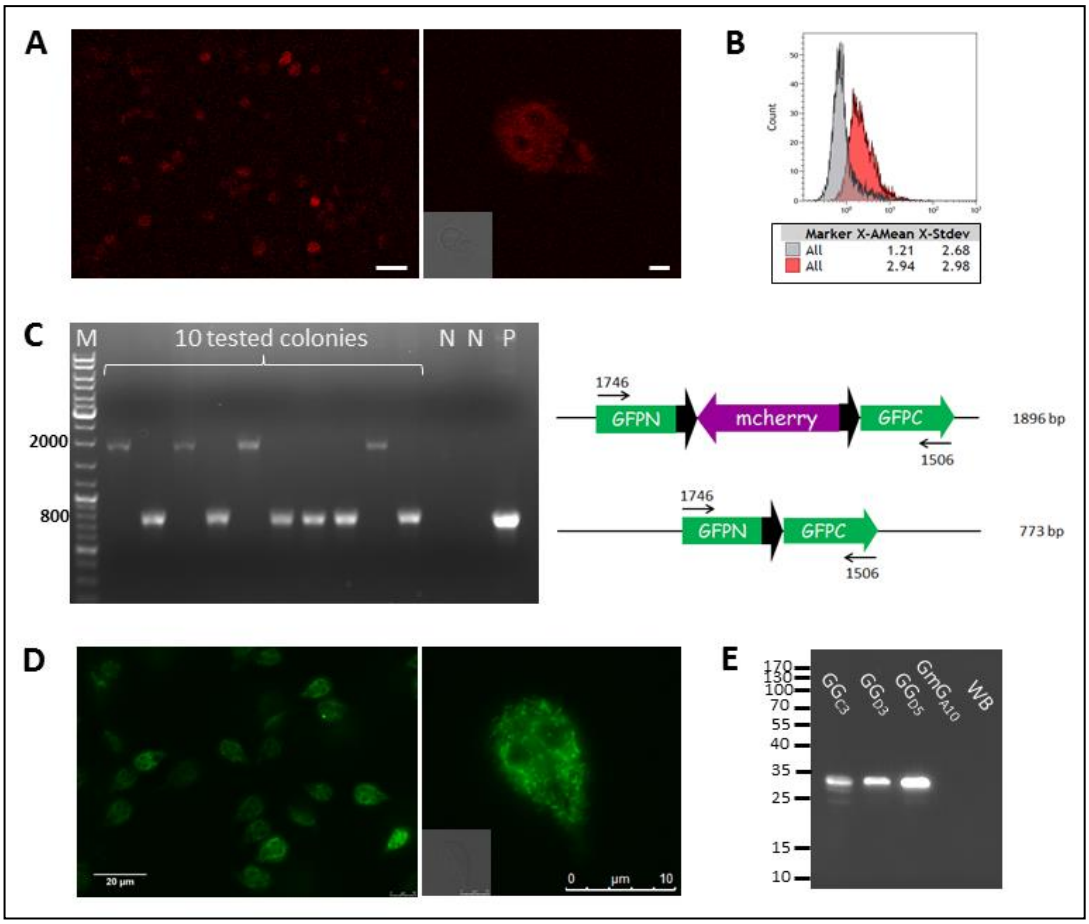


Figure 2: *Cre mediated excision reconstitutes GFP expression.* **(A)** The GFPN-mcherry-GFPC reporter line expressing mCherry prior to Cre-mediated removal of the floxed mcherry expression cassette. Live cell imaging by confocal microscopy shows red fluorescent trophozoites. Scale bars: 20 µm (left) and 2 µm (right). **(B)** Flow cytometry analysis reveals a 2.4-fold higher relative average fluorescence intensity for the mcherry reporter line (red, X-AMean = 2.94) compared to wild type trophozoites (grey, X-AMean = 1.21). *X-AMean:* arithmetic mean of the relative fluorescence intensity; *X-Stdev:* standard deviation of X-AMean. **(C)** Analysis of single cell-derived colonies obtained from the GFPN-mcherryfl-GFPC reporter line after Cre-induction and single cell sorting. Primer pair 1746 and 1506 was used to check for the retention of the floxed mcherry cassette (expected product size ~ 1900bp) or the successful reconstruction of the GFP ORF (expected product size ~780bp). 6 colonies out of 10 show a PCR product for reconstruction of the GFP ORF. *N:* negative control gDNA from wild type cells. *P:* positive control GFPN-mcherryfl-GFPC-encoding plasmid; *M:* size reference marker. **(D)** The reporter line after Cre-mediated removal of the floxed mCherry cassette. Wide field immunofluorescence microscopy depicting GFP-expressing cells. Scale bars: 20 µm (left) and 10 µm (right). **(E)** Western blot analysis confirms the presence of the GFP reporter (32 kDa) in three selected single cell-derived colonies that positive by PCR for GFP reconstitution (lines GG C3, D3, D5). The reporter is not detected in a GFPN-mcherryfl-GFPC (GmG) line that was negative by PCR for GFP reconstitution (GmG A10) nor in wild type control cells (WB). *GmG:* GFPN-mcherryfl-GFPC; *GG:* GFPN-loxP-GFPC; *M:* protein ladder.

Chromosomal excision of a drug resistance marker in a transgenic cell line using the Cre/loxP system

In order to determine if the Cre/loxP system can be implemented for the removal of a drug-resistance marker from the genome of stably transfected *G. lamblia*, we developed a reporter construct coding for a BiP-GFP expression cassette and a floxed puromycin resistance cassette (Purofl-BiP-GFP, figure 3A). The GFP reporter is under the control of the constitutive Bip promoter and targeted to the endoplasmic reticulum. We generated a *G. lamblia* cell line stably transfected with construct Purofl-BiP-GFP (figure 3A); we confirmed stable integration into the intergenic region flanking the TPI gene by PCR and DNA sequencing. Expression of the BiP-GFP reporter was verified by fluorescence microscopy analysis (not shown). Since the expression level of the GFP reporter was variable within the cells of a population, we performed single cell sorting to select for the strongest BiP-GFP expressing clonal lines (figure S2).

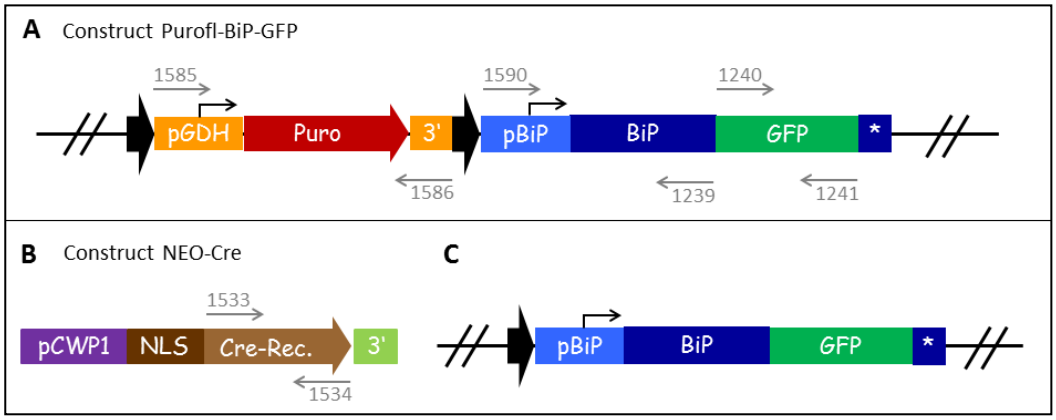


Figure 3: *Schematic illustration of the puromycin recycling strategy*. **(A)** Cells stably transfected with construct Purofl-BiP-GFP encode for the reporter protein BiP-GFP downstream of a floxed puromycin resistance cassette. Black lines represent the site of genomic integration for construct Purofl-BiP-GFP. *Puro:* puromycin N-acetyltranferase; *GDH:* glutamate dehydrogenase (GL50803_21942); *pGDH:* GDH promoter; *3' (orange):* 3' UTR of GDH; *BiP:* GlHsp70/BiP (GL80503_17121); *pBiP:* Hsp70/BiP promoter sequence; *asterisk:* KDEL encoding sequence; black arrow: loxP site; *grey arrows:* oligonucleotides used for PCR amplification of fragments (table S1). **(B)** The Cre-recombinase is encoded on a separate construct named NEO-Cre. *CWP1:* cyst wall protein 1 (GL50803_5638); *3' (green):* 3'UTR of CWP1; *pCWP1:* CWP1 promoter region, *NLS:* nuclear localization signal; *Cre-rec:* Cre-recombinase; *grey arrows:* oligonucleotides used for PCR amplification of fragments (table S1). **(C)** Loss of the floxed puromycin resistance cassette and maintenance of the BiP-GFP reporter after transient Cre expression in cell lines with integrated Purofl-BiP-GFP. Following excision, a single loxP site remains upstream of the BiP-GFP cassette. Black lines represent the site of genomic integration for construct Purofl-BiP-GFP.

As previously described, we introduced Cre into clonal Purofl-BiP-GFP cell lines by transfecting the episome NEO-Cre (figure 3B) and selecting transgenic cells for resistance to G418. Transient expression of Cre was induced by a pulse of 1.5 hours in induction medium after which cells were cultured again in standard growth medium. Preliminary analysis of the Cre-pulsed population by PCR and DNA sequencing demonstrated successful excision of the floxed puromycin resistance cassette from the genome (not shown). The cells were grown for 21 days in the absence of G418 selection to promote loss of the NEO-Cre episome. To obtain clonal BiP-GFP expressing cell lines with an excised puromycin cassette (figure 3C), we used single cell sorting (figure S3). PCR analysis of gDNA extracted from 10 single cell-derived clonal lines performed 4 weeks post induction of Cre demonstrated successful excision of Purofl from the chromosomal DNA in all lines (figure 4, middle, top panel). However, further analyses demonstrated that all clonal cell lines still contained the puromycin resistance-encoding sequence (figure 4, middle, second panel) which, to a certain extent, was located upstream of the BiP-GFP expression cassette (figure 4, middle, third panel). The Cre-encoding sequence was also amplified in all cells (figure 4, middle, fourth panel), indicating that the episomal NEO-Cre vector was still present - even after 4 weeks of cell culture without selection. This result was in line with the data presented in the previous section, where we demonstrated episomal maintenance of the floxed mCherry cassette following Cre-mediated equilibrium even in the absence of positive selection for the NEO-Cre vector (*25, 31*).

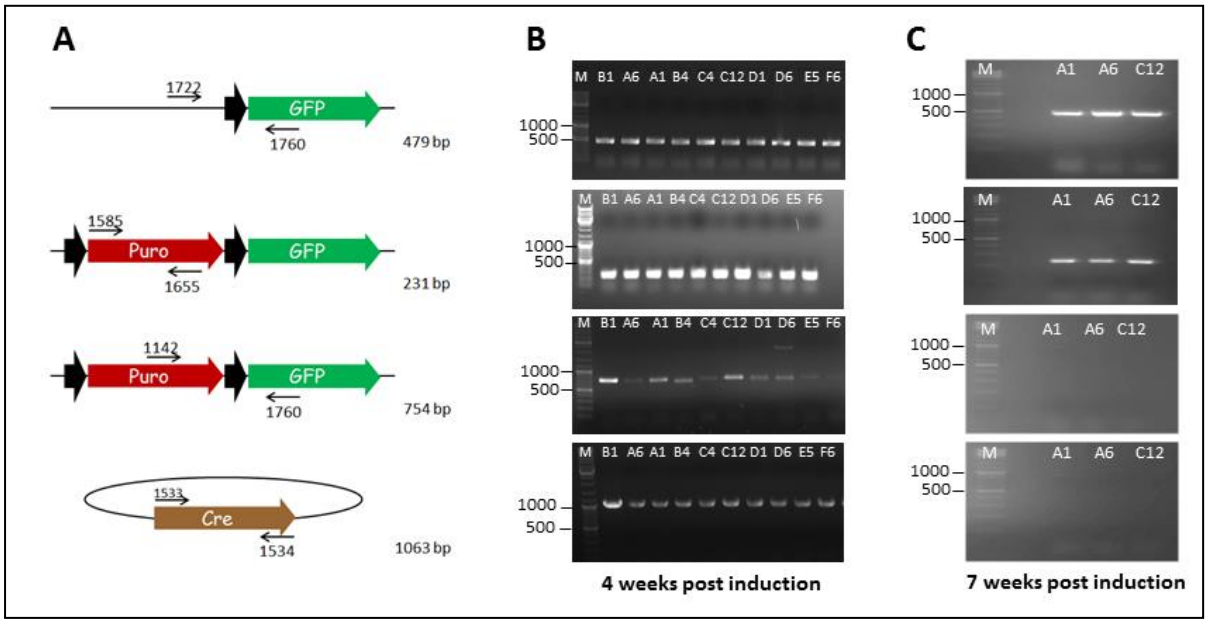


Figure 4: *Analysis of transgenic clonal lines obtained from a Purofl-BiP-GFP parent strain after Cre-induction und single cell sorting.* **(A)** Primers used for PCR analysis to check for the excision of Puro (first panel), the presence of Puro (second panel), the position of Puro upstream of the BiP-GFP expression cassette (third panel), and the presence of Cre (fourth panel). The expected product size for each PCR reaction is indicated. (**B)** PCR analysis performed at 4 weeks post induction of Cre. All ten clonal cell lines analyzed show the PCR product for excised Puro (fist panel), presence of Puro (second panel), position of Puro upstream of BiP-GFP (third panel) and presence of Cre (fourth panel). (**C)** PCR analysis performed at 7 weeks post induction of Cre. All three clonal lines analyzed (A1, A6, C12) show the PCR product for excised Puro (first panel) and for the presence of Puro (second panel). None of the three strains analyzed shows a PCR product for Puro upstream of the BiP-GFP cassette (third panel) or for the presence of Cre (fourth panel).

Three clonal cell lines, A1, A6 and C12, were selected for further analysis and cultured in standard growth medium without antibiotic selection for 3 additional weeks. A repetition of the PCR-based analysis of gDNA extracted from each line revealed a different picture. Successful excision of the floxed puromycin expression cassette (figure 4, right, first panel) as well as the persistence of the puromycin-encoding region (figure 4, right, second panel) was confirmed in all 3 lines. However, the puromycin-encoding region was not linked to the BiP-GFP expression cassette anymore (figure 4, right, third panel) and no Cre-encoding sequence was amplified (figure 4, right, fourth panel), indicating that, after several weeks of growth in non-selective culture medium, the NEO-Cre plasmid had been lost.
Interestingly, the excised, floxed puromycin cassette appeared to be maintained as a non-integrated DNA element, most likely in the form of a small circularized DNA molecule (as excised by the Cre enzyme), despite the absence of any engineered origin of replication. We tested whether this non-integrated element was able to confer resistance to puromycin by culturing lines A1, A6 and C12 in standard growth medium supplemented with 40 µg/ml puromycin, the concentration routinely used for selection of resistant cells in culture. After 3 days in selective medium, no viable cells were detected, whereas cell confluence was reached in non-selective medium (data not shown). This suggests that persistence of the circularized, episomally maintained puromycin cassette does not confer resistance to the antibiotic.

Taken together, our data demonstrate that the Cre/loxP system can be applied to remove a floxed puromycin resistance gene after integration into the *G. lamblia* genome and re-establish sensitivity to puromycin. The Cre-recombinase mediates equilibrium between excision and insertion of the floxed DNA sequence. The equilibrium, however, is driven towards excision and stagnates upon loss of the Cre-encoding plasmid after 4-7 weeks rendering cells sensitive to puromycin again.

Recycling of the transfected puromycin resistance allows the generation of dual reporter *G. lamblia* cell lines

We next tested whether we could re-utilize the puromycin selection marker in a subsequent transfection after excision of the previously stably integrated puromycin resistance cassette. To do this, we used clonal cell line A1 which constitutively expressed BiP-GFP as a parent cell line for transfection with constructs encoding either a mitosome reporter (GL50803_17161; construct PTom40HA) (*32*) or an ER exit site component (GL50803_9376; construct PSec23HA) (*12*), both tagged at the C' terminus with an HA epitope. Expression of both reporter genes is driven by their respective promoter sequences (*12*). Both vectors carry a constitutively-expressed puromycin resistance cassette for selection of antibiotic resistant transgenic parasites. Line A1 was transfected according to established protocols and transgenic cells were selected using 40 µg/ml puromycin added to standard growth medium. Resistant cells were detected after ~5 days and were harvested at confluence for analysis by immunofluorescence microscopy. While the parent A1 line expressed only BiP-GFP (figure 5, top), the puromycin-resistant transgenic lines expressed in addition HA-tagged GlSec23 (figure 5, middle) or GlTom40 (figure 5, bottom). Both reporters localized at the appropriate subcellular structures, the ER exit sites or mitosomes, in the majority of cells (*12, 33*). This demonstrates that we could indeed re-utilize the puromycin resistance gene to generate novel dual reporter *G. lamblia* lines in a transgenic parent cell line expressing an ER marker (BiP-GFP) in combination with either a COPII subunit (GlSec23) or a

mitosome component (GlTom40). The experiment also shows for the first time that antibiotic sensitive, transgenic parent cell lines can be used for further genetic manipulation and expression of additional reporter proteins without negative effects.

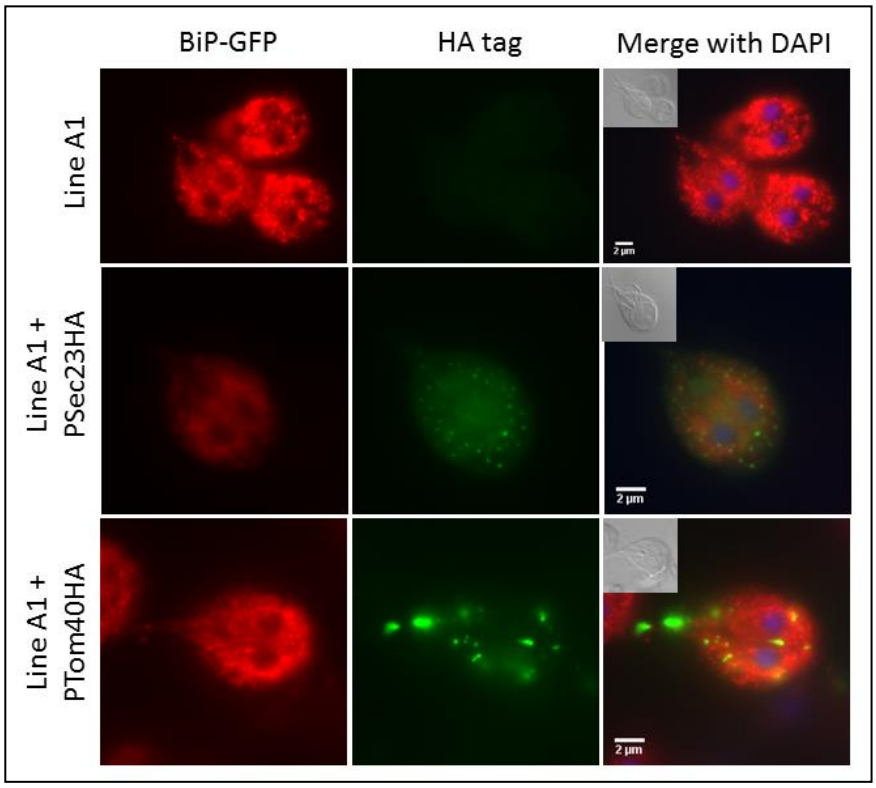


Figure 5: *The puromycin resistance cassette can be re-used for further transfection rounds.* Immunofluo-rescent antibody labelling and wide-field microscopy analysis of parent line A1 (upper panels) and secon-darily transfected line A1 with either construct PSec23HA (middle panels) or PTom40HA (bottom panels). While parent A1 cells present only BiP-GFP labelling of the ER (in red), trans-fectants with both constructs present additional HA labelling (in green) localized either to ER exit sites (middle panels) or to mitosomes (lower panels). Nuclear DNA was labelled with DAPI and is shown in the merged images. Insets: DIC images. Scale bars: 2 µm.

Taken together, our data demonstrate how, following removal of a floxed puromycin resistance cassette with a short pulse of Cre recombinase expression, we successfully "recycled" the puromycin resistance cassette for subsequent episomal secondary transfections and establishment of dually transfected Giardia cell lines.

Discussion

G. lamblia's tetraploid status has severely hampered the application of reverse genetic approaches to study molecular and cellular processes in this parasitic protozoan. Currently, suitable strategies for the study of gene function in Giardia are restricted to gene overexpression using transfection of plasmid vectors (*7, 9, 10*), and mRNA-dependent knockdown by targeting mRNA production, stability, or accessibility of the translation machinery. Several knock down approaches have been reported, including the expression of antisense-sequences (*15, 16*) or long-dsRNA sequences (*17, 18*), and the transfection with hammerhead ribozymes (*19, 20*) or morpholinos (*21*). However, inconsistent levels of down-regulation of expression have frequently been observed (*16, 18-21*). Since morpholinos are designed to bind their target mRNA in the region between the 5'cap and 25 nucleotides downstream of the translation start codon, the application of this tool in *G. lamblia* is challenging due to

the unusually short 5' untranslated regions (5'UTR) (*34*). Furthermore, targeting specificity is a challenge when dealing with gene families which may present very high levels of sequence similarity. This makes ribozyme and morpholino sequence design difficult and may confound phenotypic analysis. In addition, since morpholinos cannot be replicated, any phenotype resulting from their presence is not genetically sustainable. This implies that any ensuing analysis has to be performed within a very short time from morpholino delivery. Taken together, current strategies for loss-of-function studies in *G. lamblia* present several disadvantages and, in the best case, provide information only on the biological impact of reduced protein levels. This may not be sufficient to address questions concerning the exact role and function of abundant proteins.

To address the challenge of manipulating a tetraploid genome by physically targeting DNA sequences rather than depleting mRNA levels and/or functionality, we implemented the phage-derived Cre/loxP system (*23*) in *G. lamblia*. We demonstrated that Cre is functionally expressed in this protozoan species and that the Cre/loxP-based strategy can be successfully employed to remove a floxed puromycin resistance cassette from the *G. lamblia* genome. This allowed for "recycling" of a selectable puromycin resistance marker for a further transfection round of the same cell line.
The implications for the effective removal and re-utilization of a genomically-integrated puromycin antibiotic resistance cassette are of crucial importance to Giardia researchers for at least four reasons:

i) Given the limited number of antibiotic selection markers applicable to transgenic Giardia cultures (*11*), our Cre-based strategy permits repeated use of a single drug-resistance cassette. This opens up exciting opportunities for serial manipulations of the Giardia genome using a single selection marker. In this work, we demonstrated the feasibility of this approach by repeatedly transfecting transgenic parent lines using a single resistance marker.

ii) Previous reports have shown how the maintenance of a constitutively-expressed puromycin resistance cassette causes off-target effects which include a partial de-regulation of *G. lamblia* differentiation (*35*). Therefore, Cre-mediated removal of the antibiotic resistance cassette allows for a less confounding evaluation of any transgene-related phenotype.

iii) Taking this a step further, we can envisage the application of the Cre/loxP system and its variants (*25*) to the serial removal of any target gene, thus paving the way towards *bona fide* gene knock-out studies in *G. lamblia*, which are currently not feasible.

iv) The generally high degree of sequence divergence in Giardia makes utilization of heterologous antibodies in cell biology research, even against proteins which are otherwise highly conserved across species (*36*), highly impractical. Here we provide an alternative to generating homologous antibodies in the form of transgenic, antibiotic sensitive parent cell lines which constitutively express organelle markers and can be super-transfected.

The main limitation when applying the Cre/loxP method to genomic manipulations in *G. lamblia* concerns stringent regulation of Cre recombinase expression once it is induced. We found that, in the

absence of positive neomycin-based selection, the Cre-encoding plasmid (NEO-Cre) was still present after more than 4 weeks in cultured cell populations. Although the Cre-encoding ORF was engineered downstream of a tightly-regulated inducible promoter, the persistence of even very few Cre molecules may decrease the overall efficiency of excision over (re-)integration by maintaining equilibrium between the two phenomena (*25, 31*).

To address this challenge, Cre-recombinase could be electroporated as an enzyme directly into cells, similarly to restriction enzymes in the restriction mediated enzyme integration (REMI) technique (*37*). This would allow for a more rapid loss of recombinase activity as a result of protein turn-over and dilution during cell division. However, this may also result in an overall reduction in the efficiency of excision and would almost certainly require single cell sorting procedures to establish homogenous excised cell lines.

Alternatively, Cre recombinase activity itself could be subject to regulation. Recently, Andenmatten and colleagues developed the DiCre system in *T. gondii* as an alternative strategy for induction of Cre activity (*38*). This method is based on the conditional expression of Cre as two separate subunits which are fused to either the FK506 Binding Protein (FKBP) or the rapamycin binding domain of mTOR (FRB), respectively (*38*). Rapamycin-dependent complexing of the FKBP-FRB moieties (*39*) results in dimerization of the two Cre-subunits and full Cre recombinase activity.

In summary, by establishing the Cre/loxP system in *Giardia* and by demonstrating that this novel tool can be used to serially manipulate the parasite's genome, we have reached a new milestone in genome and cell biological research of this tetraploid protozoan. We are convinced that the application of the Cre/loxP technique provides a powerful and novel means for more complex and versatile genetic manipulations in *G. lamblia*, and we predict that this method will significantly impact investigation of gene and protein function in this evolutionary diverged parasitic protozoan.

Methods

Giardia cell culture and in vitro encystation

Giardia lamblia WBC6 (ATCC catalog number 50803) trophozoites were grown under anaerobic conditions in 11 ml culture tubes (Nunc, cat. 156758) containing TYI-S- 33 medium supplemented with 10% foetal bovine serum and bovine bile according to standard protocols (*29*). Parasites were harvested by chilling the tubes on ice for 30 minutes to detach adhering cells and collected by centrifugation (900 x *g*, 10 minutes, 4 °C).

Encystation was induced using the two-step method as described previously (*40*) by cultivating the trophozoites in bile-free medium for 44 hours and thereafter in medium (pH 7.85) containing porcine bile.

Cloning of expression vectors

Construct GFPN-mcherryfl-GFPC: We designed an expression vector encoding the N-terminal half of GFP (GFPN, 270 bp) and the C-terminal half of GFP (GFPC, 447 bp) as described previously (*41*) and positioned a floxed mCherry-encoding expression cassette (mcherryfl) in the opposite orientation in between

(figure 1A). The two loxP sites were positioned in the same orientation to permit Cre-mediated excision of mcherryfl to reconnect GFPN and GFPC, which reconstitutes a full-length ORF for GFP (figure 1C). The *G. lamblia* Hsp70/BiP (GL50803_17121) putative promoter sequence including the N-terminal 116 bp region of the ORF, encoding also the secretory signal peptide, was placed upstream of ORFs GFPC and mcherryfl to achieve constitutive and high level expression of ER-targeted fluorescent reporter proteins. The cassette was ligated into the previously described vector pPacV-Integ using *Xba*I and *Pac*I restriction sites (*30*). Oligonucleotides used for engineering of constructs are indicated in figure 1A and listed in table S1.

Construct Purofl-BiP-GFP: We used the previously described pPacV-Integ vector (*30*) and inserted two loxP sites in the same orientation on either side of the puromycin resistance (Puro) cassette. Downstream of the floxed Puro cassette, we placed an expression cassette for a Hsp70/BiP (GL50803_17121) promoter-driven C-terminal GFP fusion to the Hsp70/BiP (GL50803_17121) coding sequence. The GFP moiety was further modified at the 3' end by including a 12bp Lys-Asp-Glu-Leu coding sequence to insert an ER-retrieval signal. This resulted in a constitutively-expressed BiP-GFP fusion predicted to be subject to ER retrieval. Oligonucleotides used for construct engineering are indicated in figure 3A and listed in table S1.

Construct NEO-Cre: The Cre-recombinase encoding sequence (GenBank accession number NC_005856.1) was amplified from plasmid p116, a gift from Dr. Kurt Tobler (Institute of Veterinary Virology, University of Zurich, Switzerland). The Cre ORF was modified at the 5' end by including a 18bp Lys-Lys-Lys-Arg-Lys-Val coding sequence to insert a nuclear localization signal (*42*). To permit inducible expression of the Cre-recombinase, the promoter sequence of *G. lamblia* cyst wall protein 1 (GL50803_5638) was positioned upstream of the Cre encoding region. The construct was ligated into the previously described RAN-neo vector using *Xba*I and *Pac*I restriction sites (*10, 29*).

Constructs PSec23HA and PTom40HA: Constructs PSec23HA encodes a C-terminally HA-epitope tagged variant of the ER exit site marker GlSec23 (Gl50803_9376) which is positioned downstream of its endogenous promoter sequence and carries a constitutively-expressed puromycin expression cassette. The construct was designed as described previously (*12*). Construct PTom40HA encodes a C-terminally HA-epitope tagged variant of the protein translocation channel (GL50803_17161), a component of the *G. lamblia* translocase of the outer membrane (TOM) complex localized in the outer mitosomal membrane (*32*). The Tom40 encoding sequence and its promoter region were amplified using oligonucleotides 1295 and 1309 (listed in table S1) and the construct was inserted into the previously described vector pPacV-Integ using *Xba*I and *Pac*I restriction sites (*30*).

Plasmid transfection

pPacV-Integ based plasmids prepared from *E.coli* were linearized using *Swa*1 restriction endonuclease (*16*), and 15 µg of digested DNA was electroporated (BIO RAD Gene Pulser, 350V, 960 mF, 800 Ohm) into 3 x 10^e6 freshly harvested trophozoites. Linearized plasmids were targeted to the *G. lamblia* triose phosphate isomerase (GlTPI) locus and integration occurred by homologous recombination under

selective pressure of the antibiotic puromycin (Invitrogen, cat. ant-pr-1) at a concentration of 40 µg/ml. Following selection of resistant clones, transgenic cell lines were maintained and analyzed without antibiotic. Stable transfection was confirmed by PCR on extracted genomic DNA (gDNA) resulting in ~2000 bp amplicons spanning the predicted GI-TPI integration site for pPacV-Integ-derived constructs using the primer pair 1440 and 1441 (*12*). gDNA quality was assessed by PCR performed in parallel using primers 1460 and 1461, resulting in a 999 bp product which corresponds to the GL50803_9157 gene with its promoter sequence. Oligonucleotide sequences are listed in supplementary table S1.
For episomal maintenance of the NEO-Cre plasmid, trophozoites were transfected with circular expression vectors which do not integrate into the genome of the WB isolate and grown under selective pressure with neomycin (Sigma, cat. G8168) at a concentration of 140 µg/ml.

Oxidative GFP chromophore formation and single cell sorting

5 x 10e7 mCherry or GFP-expressing trophozoites cells were harvested as described, resuspended in 1.5 ml standard growth medium and transferred to 24 well plates (Sigma, cat. Z707791). Following incubation on ice for 16 h, oxygenated cells were harvested and protein folding was completed by incubation for 30 minutes at 37°C. The samples were washed once (900 x *g*) in PBS supplemented with 5 mM glucose (Fluka, cat. 49139), 5 mM L-cysteine (Sigma, cat. C6852), and 0.1 mM ascorbic acid (Fluka, cat. 95209) at pH 7.1. The cell pellet was resuspended in 200 ul of supplemented PBS.
Single cell sorting was performed on a BD FACSAriaIII™ cell sorter. For data acquisition and processing, BD FACSDiva™ software (version 6.3.1) was used. The sort was performed using a nozzle with a 100 micrometer orifice diameter at 20 psi sheath pressure. GFP and mCherry were excited by a 488 nm and a 561 nm laser, respectively, and emission was detected using 510/30 and 616/23 band pass filters. The cell populations were defined by a parent gate P1 based on FSC-A and SSC-A. An additional gate P2 based on FSC-A and FSC-H was set to select for single events. GFP and mCherry positive cells were defined by gate P3 in the respective fluorescence channels. Wild type trophozoites were used as negative control. The cells were sorted into 96 well plates (Sigma, cat. Z707902) containing Giardia standard growth medium. Sorted cells were grown in anaerobic conditions in the presence of reducing agents (BD GasPak™, cat. 260683). After 5 days, single cell derived clonal populations were transferred to 11 ml culture tubes (Nunc, cat. 156758) for further culturing and analysis.

Transient activation of Cre-recombinase expression:

After single cell sorting, strains *Purofl-BiP-GFP* and *GFPN-mcherryfl-GFPC* were transfected with the episomally maintained *NEO-Cre* vector. Transient activation of Cre-recombinase expression under the control of the tightly regulated, encystation-specific cyst wall protein 1 (CWP1) promoter was achieved by incubating trophozoites in encystation medium for 1.5 h (*40*) before returning them to standard culture conditions. This interval is sufficient to generate a brief pulse of Cre-recombinase expression but not long enough for trophozoites to commit to encystation (*40*).

Immunolabelling assay

Chemical fixation and preparation for fluorescence microscopy was performed as described (*3*). Briefly, cells were washed with cold PBS after harvesting and fixed with 3% formaldehyde in PBS for 40 min at

20°C, followed by 5 minutes incubation with 0.1 M glycine in PBS. Cells were permeabilized with 0.2% Triton X-100 in PBS for 20 min at room temperature and blocked overnight in 2% BSA in PBS. Incubations of all antibodies were done in 2% BSA/0.2% Triton X-100 in PBS for 1 h at 4°C. The following antibodies were used in this work: FITC-conjugated anti-HA epitope (Roche Diagnostics GmbH, Manheim, Germany; dilution 1:50), Alexa594-conjugated anti-mouse (Molecular Probes, USA; dilution 1:300) and anti-GFP monoclonal (Clontech; dilution 1:300). Post incubation washes were done with 1% BSA/0.1% triton X-100 in PBS. Labeled cells were embedded for microscopy with Vectashield (Vector Laboratories, Inc., Burlingame, CA, USA) containing the DNA intercalating agent 4'-6-Diamidino-2-phenylindole (DAPI) for detection of nuclear DNA.

Microscopy and flow cytometry-based analysis

Fluorescence analysis was performed either on a wide-field fluorescence microscope (Leica DMI 6000B) with LAS AF software (version 3.1.0) for data collection, or on a Leica SP2 AOBS confocal laser-scanning microscope (Leica Microsystems, Wetzlar, Germany) equipped with a glycerol objective (Leica, HCX PL APO CS, 63x, 1.3 Corr) and Leica Confocal software (version 2.61) for data collection. WCIF ImageJ (version 1.37c) was used for image processing.
Flow cytometry-based analysis was performed on a BC Gallios^{TM} flow cytometer with Gallios Software for data acquisition. GFP and mCherry were excited by a 488 nm and an external 561 nm laser, and emission was detected by 525/40 and a 620/30 band pass filters, respectively. Kaluza software (version 1.2) was used for data processing.

Acknowledgements:

We thank Therese Michel for excellent technical assistance, and the team of the Flow Cytometry Facility of the University of Zurich for their support in single cell sorting.

References

1. A. G. Simpson, A. J. Roger, The real 'kingdoms' of eukaryotes. *Curr Biol* **14**, R693 (Sep 7, 2004).
2. H. G. Morrison *et al.*, Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science* **317**, 1921 (Sep 28, 2007).
3. M. Marti *et al.*, An ancestral secretory apparatus in the protozoan parasite Giardia intestinalis. *J Biol Chem* **278**, 24837 (Jul 4, 2003).
4. J. B. Dacks, G. Walker, M. C. Field, Implications of the new eukaryotic systematics for parasitologists. *Parasitol Int* **57**, 97 (Jun, 2008).
5. K. S. Kabnick, D. A. Peattie, In situ analyses reveal that the two nuclei of Giardia lamblia are equivalent. *J Cell Sci* **95 (Pt 3)**, 353 (Mar, 1990).
6. R. Bernander, J. E. Palm, S. G. Svard, Genome ploidy in different stages of the Giardia lamblia life cycle. *Cell Microbiol* **3**, 55 (Jan, 2001).
7. S. M. Singer, J. Yee, T. E. Nash, Episomal and integrated maintenance of foreign DNA in Giardia lamblia. *Mol Biochem Parasitol* **92**, 59 (Apr 1, 1998).

8. M. K. Poxleitner *et al.*, Evidence for karyogamy and exchange of genetic material in the binucleate intestinal parasite Giardia intestinalis. *Science* **319**, 1530 (Mar 14, 2008).
9. D. C. Yu, A. L. Wang, C. C. Wang, Stable coexpression of a drug-resistance gene and a heterologous gene in an ancient parasitic protozoan Giardia lamblia. *Mol Biochem Parasitol* **83**, 81 (Dec 2, 1996).
10. C. H. Sun, C. F. Chou, J. H. Tai, Stable DNA transfection of the primitive protozoan pathogen Giardia lamblia. *Mol Biochem Parasitol* **92**, 123 (Apr 1, 1998).
11. S. R. Davis-Hayman, T. E. Nash, Genetic manipulation of Giardia lamblia. *Mol Biochem Parasitol* **122**, 1 (Jun, 2002).
12. C. Faso, C. Konrad, E. M. Schraner, A. B. Hehl, Export of cyst wall material and Golgi organelle neogenesis in Giardia lamblia depend on endoplasmic reticulum exit sites. *Cell Microbiol*, (Oct 25, 2012).
13. I. J. Macrae *et al.*, Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195 (Jan 13, 2006).
14. C. G. Prucca *et al.*, Antigenic variation in Giardia lamblia is regulated by RNA interference. *Nature* **456**, 750 (Dec 11, 2008).
15. M. C. Touz, J. T. Conrad, T. E. Nash, A novel palmitoyl acyl transferase controls surface protein palmitoylation and cytotoxicity in Giardia lamblia. *Mol Microbiol* **58**, 999 (Nov, 2005).
16. S. Stefanic *et al.*, Neogenesis and maturation of transient Golgi-like cisternae in a simple eukaryote. *J Cell Sci* **122**, 2846 (Aug 15, 2009).
17. M. C. Touz, L. Kulakova, T. E. Nash, Adaptor protein complex 1 mediates the transport of lysosomal proteins from a Golgi-like organelle to peripheral vacuoles in the primitive eukaryote Giardia lamblia. *Mol Biol Cell* **15**, 3053 (Jul, 2004).
18. M. R. Rivero, L. Kulakova, M. C. Touz, Long double-stranded RNA produces specific gene downregulation in Giardia lamblia. *J Parasitol* **96**, 815 (Aug, 2010).
19. M. Dan, A. L. Wang, C. C. Wang, Inhibition of pyruvate-ferredoxin oxidoreductase gene expression in Giardia lamblia by a virus-mediated hammerhead ribozyme. *Mol Microbiol* **36**, 447 (Apr, 2000).
20. A. Castillo-Romero *et al.*, Rab11 and actin cytoskeleton participate in Giardia lamblia encystation, guiding the specific vesicles to the cyst wall. *PLoS Negl Trop Dis* **4**, e697 (2010).
21. M. L. Carpenter, W. Z. Cande, Using morpholinos for gene knockdown in Giardia intestinalis. *Eukaryot Cell* **8**, 916 (Jun, 2009).
22. L. Li, C. C. Wang, Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, Giardia lamblia. *J Biol Chem* **279**, 14656 (Apr 9, 2004).
23. N. Sternberg, D. Hamilton, R. Hoess, Bacteriophage P1 site-specific recombination. II. Recombination between loxP and the bacterial chromosome. *J Mol Biol* **150**, 487 (Aug 25, 1981).
24. R. H. Hoess, M. Ziese, N. Sternberg, P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc Natl Acad Sci U S A* **79**, 3398 (Jun, 1982).
25. C. S. Branda, S. M. Dymecki, Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* **6**, 7 (Jan, 2004).

26. M. T. O'Neill, T. Phuong, J. Healer, D. Richard, A. F. Cowman, Gene deletion from Plasmodium falciparum using FLP and Cre recombinases: implications for applied site-specific recombination. *Int J Parasitol* **41**, 117 (Jan, 2011).
27. S. Brecht, H. Erdhart, M. Soete, D. Soldati, Genome engineering of Toxoplasma gondii using the site-specific recombinase Cre. *Gene* **234**, 239 (Jul 8, 1999).
28. B. Barrett, D. J. LaCount, J. E. Donelson, Trypanosoma brucei: a first-generation CRE-loxP site-specific recombination system. *Exp Parasitol* **106**, 37 (Jan-Feb, 2004).
29. A. B. Hehl, M. Marti, P. Kohler, Stage-specific expression and targeting of cyst wall protein-green fluorescent protein chimeras in Giardia. *Mol Biol Cell* **11**, 1789 (May, 2000).
30. L. F. Jimenez-Garcia *et al.*, Identification of nucleoli in the early branching protist Giardia duodenalis. *Int J Parasitol* **38**, 1297 (Sep, 2008).
31. R. Kuhn, R. M. Torres, Cre/loxP recombination system and gene targeting. *Methods Mol Biol* **180**, 175 (2002).
32. M. J. Dagley *et al.*, The protein import channel in the outer mitosomal membrane of Giardia intestinalis. *Mol Biol Evol* **26**, 1941 (Sep, 2009).
33. A. Regoes *et al.*, Protein import, replication, and inheritance of a vestigial mitochondrion. *J Biol Chem* **280**, 30557 (Aug 26, 2005).
34. R. D. Adam, The Giardia lamblia genome. *Int J Parasitol* **30**, 475 (Apr 10, 2000).
35. L. H. Su, G. A. Lee, Y. C. Huang, Y. H. Chen, C. H. Sun, Neomycin and puromycin affect gene expression in Giardia lamblia stable transfection. *Mol Biochem Parasitol* **156**, 124 (Dec, 2007).
36. H. D. Lujan, M. R. Mowatt, J. T. Conrad, T. E. Nash, Increased expression of the molecular chaperone BiP/GRP78 during the differentiation of a primitive eukaryote. *Biol Cell* **86**, 11 (1996).
37. R. J. Wall, New gene transfer methods. *Theriogenology* **57**, 189 (Jan 1, 2002).
38. N. Andenmatten *et al.*, Conditional genome engineering in Toxoplasma gondii uncovers alternative invasion mechanisms. *Nat Methods* **10**, 125 (Feb, 2013).
39. L. A. Banaszynski, C. W. Liu, T. J. Wandless, Characterization of the FKBP.rapamycin.FRB ternary complex. *J Am Chem Soc* **127**, 4715 (Apr 6, 2005).
40. L. Morf *et al.*, The transcriptional response to encystation stimuli in Giardia lamblia is restricted to a small set of genes. *Eukaryot Cell* **9**, 1566 (Oct, 2010).
41. H. Zong, J. S. Espinosa, H. H. Su, M. D. Muzumdar, L. Luo, Mosaic analysis with double markers in mice. *Cell* **121**, 479 (May 6, 2005).
42. C. H. Sun, L. H. Su, F. D. Gillin, Influence of 5' sequences on expression of the Tet repressor in Giardia lamblia. *Mol Biochem Parasitol* **142**, 1 (Jul, 2005).

Supplementary material

Figure S1: Single cell sorting of line GFPN-mcherry-GFPC after induction of Cre. Line GFPN-mcherry-GFPC was subjected to single cell sorting after induction of Cre **(A)**. Wild type cells were used as a negative control **(B)**. The cell population is defined by parent gate P1 (red, FSC-A/SSC-A scatter plot). Gate P2 (green, FSC-A/FSC-H scatter plot) selects for single events. GFP-positive cells were selected by gate P3 (blue, SSC-A/AlexaFluor 488 scatter plot).

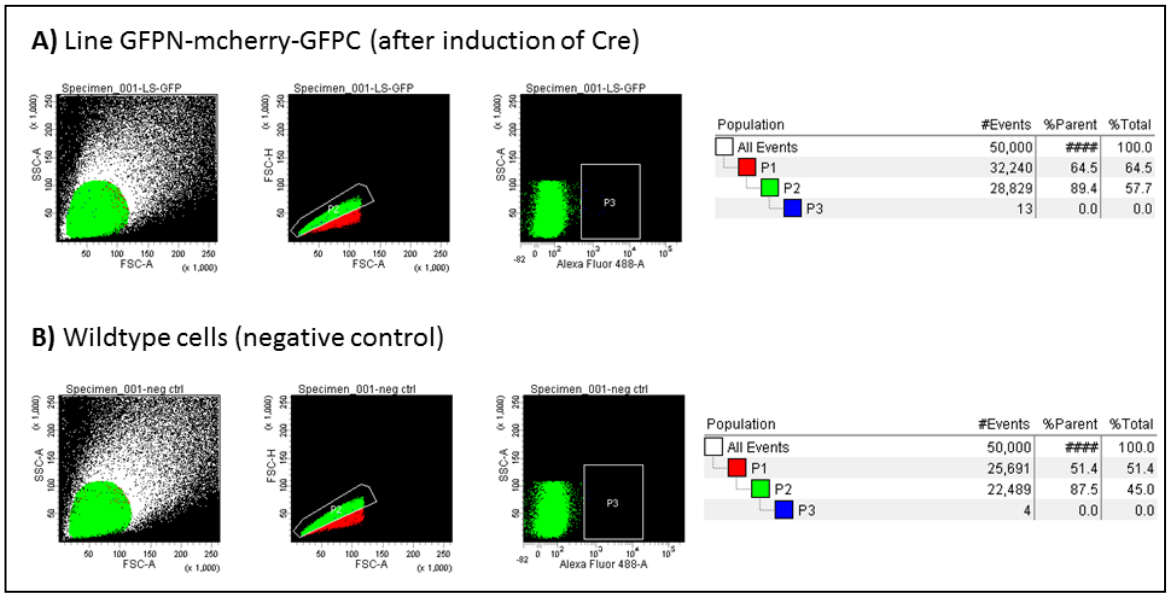


Figure S2: Single cell sorting of line Purofl-BiP-GFP before induction of Cre. Line Purofl-BiP-GFP was subjected to single cell sorting before induction of Cre **(A)**. Wild type cells were used as negative control **(B)**. The cell population is defined by parent gate P1 (red, FSC-A/SSC-A scatter plot). Gate P2 (green, FSC-A/FSC-H scatter plot) selects for single events. GFP-positive cells were selected by gate P3 (blue, SSC-A/AlexaFluor 488 scatter plot).

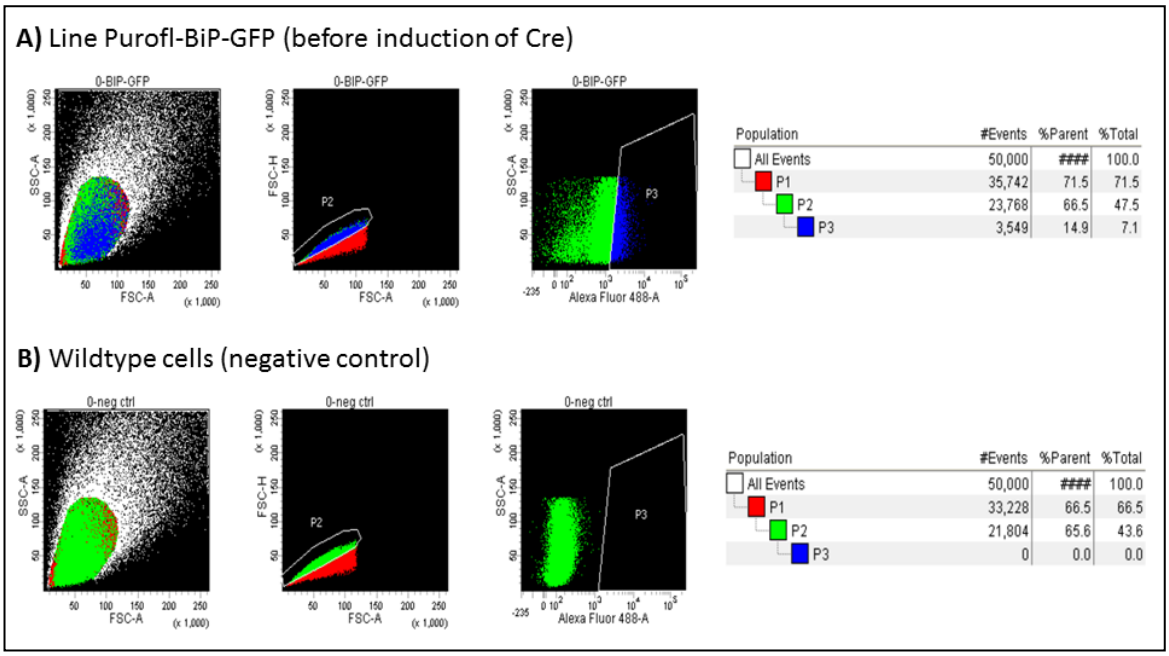


Figure S3: Single cell sorting of line Purofl-BiP-GFP after induction of Cre. Line Purofl-BiP-GFP was subjected to single cell sorting after induction of Cre **(A)**. Wild type cells were used as negative control **(B)**. The cell population is defined by parent gate P1 (red, FSC-A/SSC-A scatter plot). Gate P2 (green, FSC-A/FSC-H scatter plot) selects for single events. GFP-positive cells were selected by gate P3 (blue, SSC-A/AlexaFluor 488 scatter plot).

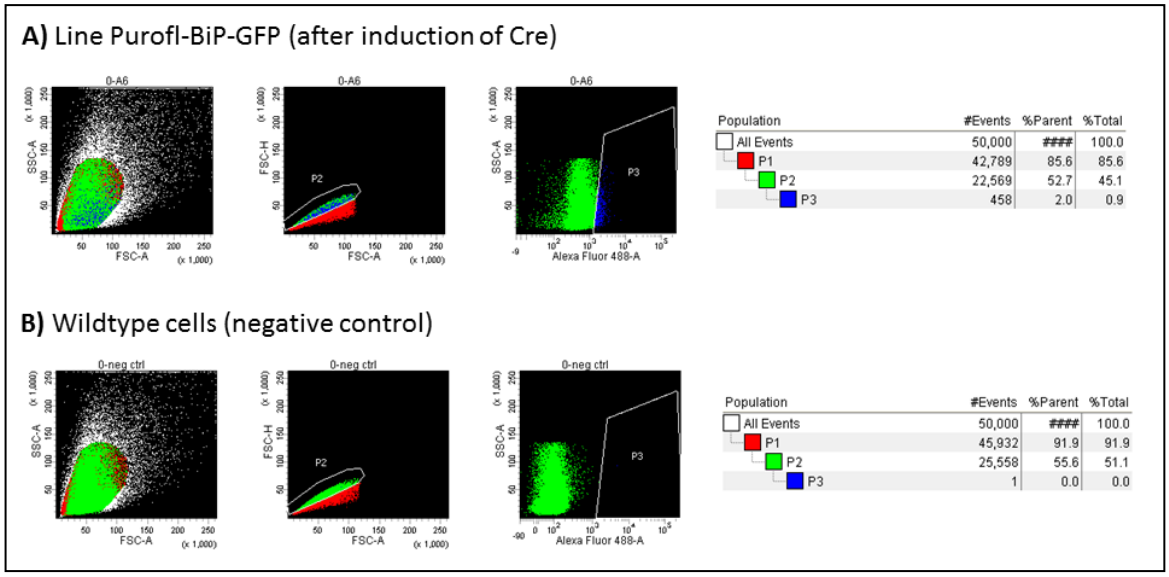


Table S1: Oligonucleotides used for construct engineering. The table includes oligonucleotide numbers (column 1), names (column 2), purpose (column 3), and sequence in 5' to 3' orientation (column 4). Restriction sites are indicated in bold face characters. Schematics for constructs and primer annealing positions are indicated in figures 1 to 4.

Primer number	Primer name	Used for	Primer sequence (5' to 3' direction)
936	CWP3p_Nhe_s	Amplification of pCWP3-CWP3 (for construct CWP3-mcherryfl-GFP)	GA**GCTAGC**GAGATAGGAGAATACTACCTG
1142	Puro_seq_fw	Check position of Puro upstream of BiP-GFP	GACCACCAGGGCAAGGGTC
1239	17121NSirev	Amplification of pBiP-BiP (for construct Purofl-BiP-GFP)	GG**ATGCAT**TTCTGCATAGTCGTACTCAGG
1240	Sbf1GFPfw	Amplification of GFP (for construct Purofl-BiP-GFP)	AC**CCTGCAGG**CCCATAAAGGAGAAGAACT TTTCAC
1241	GFP_KDELpacrev	Amplification of GFP (for construct Purofl-BiP-GFP)	CG**TTAATTAA**TTAGAGTTCATCTTTTTTGTA TAGTTCATCCATGCCA
1295	Tom40_HA_pac_rev	Amplification of pendo-Tom40HA (for construct PTom40HA)	CG**TTAATTAA**TTACGCGTAGTCTGGGACAT CGTATGGGTA CTTGTGATTGCTCGAA
1309	Tom40_promo_s_xbaI	Amplification of pendo-Tom40HA (for construct PTom40HA)	CG**TCTAGA**GCCGCAGCCTCCACGATC
1440	Integ.test for	Confirmation of stable transfection	CAAGATACTACACTCGTTGTTAATGTG
1441	Integ.test rev	Confirmation of stable transfection	GGGCTTGTACTCGGTGCCCAT
1460	p9157_XbaI_s	Quality control of genomic DNA	CG**TCTAGA**GATCAGCGCGTCTAATGGCGT G
1461	9157_SphI_as	Quality control of genomic DNA	GA**GCATGC**TGCGGATACTTCGTAGTTCCGT AG

Primer number	Primer name	Used for	Primer sequence (5' to 3' direction)
1476	GFPN_SbfI_s	Amplification of GFP N-terminal part	GA**CCTGCAGG**ATGCATAAAGGAGAAGAACTTTTCAC
1477	GFPN_loxP_as	Amplification of GFP N-terminal part	ATAACTTCGTATAATGTATGCTATACGAAGTTATCTCAGGCATGGCGCTCTTG
1506	GFP_PacI_as_longer	Amplification of GFP C-terminal part	CC**TTAATTAA**CTATTTGTATAGTTCATCCATGCCATG
1533	Cre_NLS_NsiI_s	Amplification of Cre, check for presence of Cre	GA**ATGCAT**ATGAAGAAGAAGCGCAAGGTGTCCAATTTACTGACCGTAC
1534	Cre_PacI_as	Amplification of Cre, check for presence of Cre	CG**TTAATTAA**CTAATCGCCATCTTCCAGCAG
1544	17127fwquant		CACTCTAACGTGCTGACAGT
1556	GFPC_loxP_NsiI_s	Amplification of GFP C-terminal part	CG**ATGCAT**ATAACTTCGTATAGCATACATTATACGAAGTTATTAGGATACGTGCAGGAGAGGAC
1585	GDH5UTR_NsiI_s	Amplification of the puromycin resistance cassette from a previously used plasmid (Stefanic et al., 2009), check for presence of Puro	CG**ATGCAT**GAAGCGCTGACCACAAATAAC
1586	GDH3UTR_XbaI_loxP_as	Amplification of the puromycin resistance cassette from a previously used plasmid (Stefanic et al., 2009)	CG**TCTAGA**ATAACTTCGTATAATGTATGCTATACGAAGTTATGAATTCGAGCTCGGTACCAG
1590	p17121_XbaI_s	Amplification of pBiP-BiP (for construct Purofl-BiP-GFP)	CG**TCTAGA**GACGGTCGTCCCTATCGTAATG
1602	3'UTR_XbaI_SphI_as	Amplification of mcherry	CG**TCTAGA**GCATGCCGTTTGTTTACTCACTCGTG
1604	pBiP_NsiI_s	Amplification of pBiP-LS	CG**ATGCAT**GTCGTCCCTATCGTAATGCAG
1605	BiP_LS_Afl2_as	Amplification of pBiP-LS	GT**CTTAAG**GTTGGGGATGATCTCCACTTG
1606	mcherry_Afl2_s	Amplification of mcherry	GT**CTTAAG**ATGGTGAGCAAGGGCGAGGAG
1655	Puro_seq_as	Check for presence of Puro	GTGAGGAAGAGTTCTTGCAG
1722	RS1_s	Check for excision of Puro	CTGCAGTTGCATTCGCTACAATGCTG
1760	BiP-LS_AvrII_as	Check for excision of Puro, check position of Puro upstream of BiP-GFP	GC**CCTAGG**GTTGGGGATGATCTCCACTTG
1799	CWP3_loxP_SphI_as	Amplification of pCWP3-CWP3 (for construct CWP3-mcherryfl-GFP)	GA**GCATGC**TAATAACTTCGTATAATGTATGCTATACGAAGTTATTCTGTAGTAGGGCGGCTGTATC
1800	mcherry_SphI_s	Amplification of mcherry (for construct CWP3-mcherryfl-GFP)	GA**GCATGC**GTGAGCAAGGGCGAGGAGGATAAC
1801	CWP-3'UTR_NruI_as	Amplification of mcherry (for construct CWP3-mcherryfl-GFP)	CG**TCGCGA**TGTAAATGCATGAACTGTTCAAAGAGCCGTTTGTTTACTCACTCGTGGCTAC
1802	GFP_loxP_NruI_s	Amplification of GFP (for construct CWP3-mcherryfl-GFP)	CG**TCGCGA**ATAACTTCGTATAGCATACATTATACGAAGTTATTACATAAAGGAGAAGAACTTTTCACTG
1803	GFP_PacI_as	Amplification of GFP (for construct CWP3-mcherryfl-GFP)	CG**TTAATTAA**CTATTTGTATAGTTCATCCATGCCATG

PART V DISCUSSION AND PERSPECTIVES

1. Discussion

1.1. General

Encystation-specific vesicles (ESVs) are key organelles in the *G. lamblia* life cycle. The stage-specifically induced secretory organelles are responsible for accumulation, maturation, sorting and regulated secretion of the *Giardia* cyst wall material (CWM). The CWM (protein and carbohydrate) is secreted across the plasma membrane and deposited on the parasites surface where it forms a protective composite biopolymer that confers resistance to harsh environmental conditions. Being a critical component for *Giardia*'s survival before infection of the next host, ESVs and associated proteins constitute potential drug targets. However, only little information about their structure, composition, neogenesis and functionality is available so far.

Due to *Giardia*'s extensive ER [1] previous cell fractionation attempts to identify organelle-specific proteins have been hampered by the consistent contamination of mass spectrometry datasets with ER-derived proteins [2, 3]. To improve the specificity of the method, i.e. to increase the identification of organelle-specific hits, we attempted a conceptually new approach: Using flow cytometry we simultaneously sorted fluorescently labeled ESVs and a second set of organelles, the peripheral vesicles (PVs), using a mixed microsome fraction as starting material. Since simultaneous enrichment entails the same contaminating protein background in both fractions, we removed this large majority of hits from the mass spectrometry datasets by an *in silico* filtration procedure (Wampfler and Hehl, under review).

Our global aims were to i) develop this novel approach and obtain an ESV proteome dataset enriched in organelle-specific factors, ii) identify new ESV-associated proteins which might help us to iii) gain novel insight into the genesis, composition and functionality of these unusual secretory organelles.

Using a novel, combined approach we identified and removed a tremendous number of proteins (1059, i.e. ~90% of the combined datasets) which were common to both, ESV and PV organelle fractions. Analysis of the eliminated hits revealed a massive proportion of contaminating proteins, the majority being chaperones and factors from the ribosome/translation machinery. Based on previous attempts [2, 3], this result was expected and in line with our idea to increase the specificity of ESV and PV organelle datasets by subtraction of overlapping hits.

By removing these contaminants, we obtained a set of 72 ESV organelle and 82 PV organelle candidates. Surprisingly, partial validation of the ESV-organelle dataset by subcellular localization of 16 tagged candidates revealed many ER-derived proteins rather than novel ESV-specific factors: 11 proteins (69%) showed distribution consistent with ER, and 5 thereof (31%) were also detected in ESVs. Of the five proteins localizing to both, the ER and ESVs, none was exported to the surface of the cell during the deposition of the cyst wall. In contrast, partial validation of the PV-organelle dataset by subcellular

localization studies of 8 candidates uncovered 3 (38%) clearly PV-associated proteins and two (25%) additional PV proteins which also have secondary localization.

The main conclusions we draw from this project were that **i)** our novel combined strategy is a powerful instrument to discover organelle-specific factors but seems to strongly depend on clean separation of target organelles from their subcellular context during mechanical fractionation of the cells; **ii)** A strong physical connection between ESVs and the ER exists, whereas PVs seem to maintain a more discrete compartment identity and have a distinct set of organelle-specific proteins; **iii)** In line with previous transcriptomic analysis of encysting trophozoites [4], CWP1-3 are most likely the only highly abundant and stage-specifically expressed cyst wall proteins secreted via ESVs; and **iv)** ESV neogenesis and maturation appears to be driven primarily by inherent properties of the cargo proteins themselves with the assistance of few co-factors, if any. Our conclusions are discussed in more detail hereinafter.

1.2. Organelle identities: The critical factor for the success of the dual sorting approach

The most crucial steps in organelle proteomics are cell disruption and generation of a mixed microsome fraction (a compromise between efficient liberation of organelles from the cellular context and limitation of damage to the membrane bounded organelle), followed by the purification or strong enrichment of the target organelle by a subcellular fractionation procedure. Frequently used techniques are affinity-based isolation, density gradient centrifugation, free flow electrophoresis, and more recently also flow-cytometry [5-8]. However, despite considerable efforts to optimize purification protocols, proteins of subcellular origin distinct from the target organelle are persistent contaminants in proteomic datasets [2, 3]. In order to overcome these limitations, various *in-silico* strategies have been developed that allow the elimination of many false-positive hits from organelle proteomic datasets [3, 9].

In our study, positively sorted droplets formed by the nozzle in the flow cytometer contain one fluorescently labeled organelle. However, the organelle is basically suspended in a liquid containing a vast array of cell debris including resealed membrane microsomes, membrane sheets, and other small debris as well as soluble proteins. Simultaneous sorting of two differentially labeled organelles from this same mixed microsome fraction allowed us to identify these contaminating proteins by mass spectrometry analysis and subtract them from the organelle datasets (Wampfler and Hehl, under review).

In line with this idea, the vast majority of proteins was detected in both organelle-enriched samples and was enriched in factors associated with ribosomes and the protein translation machinery and highly abundant ER chaperones. However, the subsequent partial validation of the two organelle datasets by subcellular localization of candidates using fluorescence microscopy revealed an unexpected result: whereas several new PV-associated proteins could be confirmed, most of the candidate ESV proteins localized to the ER.

This result is most likely explained by the different subcellular contexts of the two organelles. PV organelles localize to the cortical layer just below the plasma membrane into which the ER network does not extend [10], which presumably makes separation from their subcellular context more efficient during cell disruption. In contrast, the physical separation of ESVs from their subcellular context might be severely hampered by their close connection with the ER, resulting in co-purification of ER-derived membrane fragments and associated proteins during ESV-organelle sorting. The close association of the ER and ESVs begins when ESVs are synthesized *de novo* at ER exit sites and obtain a massive amount of cargo from the ER [11, 12]. The two organelles exchange proteins in a COPI and COPII dependent manner [2, 11] and are probably even connected via an extensive network of tubular membranes [11]. Electron microscopy data clearly demonstrate that membrane continuities between the ER and even mature ESVs exist (Wampfler and Hehl, under review, [13-15]). Consistent with this, proteasomes and ribosomes are not only found at ER- but also at ESV membranes (Wampfler and Hehl, under review, [2]).

Our data demonstrate that the simultaneous organelle sorting approach is a useful approach for the discovery of novel organelle-specific proteins but strongly depends on the subcellular context of the target organelles. In addition, our data provide additional evidence for a strong physical connection between the ER and ESVs.

1.3. CWP1-3 comprise the only major, stage-specifically expressed cyst wall components

A specific, organelle-defining protein localizing exclusively to ESV at 13 hours post induction (p.i.) of encystation presupposes that its expression in trophozoites is basically zero but induced during encystation. Thus far, only the three cyst wall proteins (CWP1-3) fulfill these criteria.

A microarray study on the transcriptional regulation of gene expression in encysting *Giardia* trophozoites revealed that not more than 18 genes are significantly up-regulated during the early phase of encystation, i.e. the first 7 hours [4]. The small gene set includes factors known to play a role in encystation, such as CWP1-3, a transcription factor termed Myb1-like protein [4], and several enzymes required for the synthesis of the unique cyst wall glycan [16, 17]. Another study investigated the transcriptional response during the entire *Giardia* life cycle using serial analysis of gene expression (SAGE) [18]. A group of 42 genes was found to be upregulated during the entire process of encystation, i.e. >24h after induction. The two studies differed significantly in the mode of induction of encystation and other parameters but nevertheless identified 8 genes in common, including the CWPs and factors required for the cyst wall sugar synthesis. The small number of transcriptionally regulated genes reported by these studies suggests that with the exception of massively regulated CWPs, only relatively small adjustments in mRNA expression are required for encystation.

Comparison of the 72 candidate ESV organelle proteins identified in this proteomics study with the upregulated genes revealed only one candidate in common: the synaptic glycoprotein 2. We detected this protein of unknown function in the ER at 13 hours p.i. (Wampfler and Hehl, under review).

Although some encystation-specific factors are clearly regulated on the level of mRNA abundance, post-transcriptional regulation of gene expression seems to play an important role as well [19]. An example is an EGF-like cyst protein 1 (EGFCP1) which is trafficked, together with CWPs, via ESVs to the surface of the cell [20]. Surprisingly, the EGFCP1 transcript level significantly decreases during encystation, whilst the protein level increases, suggesting changes in translation rate, protein half-life or protein turnover. In addition to this one example, a recent large-scale proteomic study uncovered several hundred proteins that are differentially expressed during encystation, with the major changes occurring at early time points [19]. Compared to the trophozoite stage, 362 proteins were more abundant at 4 hours post induction (analysis not shown). Between 4 and 8 hours p.i., 85 proteins were differentially regulated. The lowest change was observed between 8 and 12 hours p.i. with only 7 upregulated proteins. Although this was a study attempting to identify the whole proteome of encysting cells, a comparison of the 72 candidate ESV proteins with these post-transcriptionally upregulated genes revealed 10 candidates in common, including 3 hypothetical proteins, a ribosomal protein, the dynein heavy chain, a putative importin alpha subunit, a nucleotide-triphosphate hydrolase (NTPase), an Hsp70-binding protein, a phosphatase, and a calcium-binding protein. We further investigated the calcium-binding protein and found it with a cytosolic distribution at 13 hours p.i. in fluorescence microscopy experiments (Wampfler and Hehl, under review).

Based on our finding that none of the remaining candidate ESV proteins were transcriptionally induced during encystation [4, 18] we did not expect to find any stage-specifically expressed proteins localizing exclusively to ESVs among them. However, we cannot exclude the existence of low abundance, stage-specifically regulated ESV-proteins that are not contained in our proteomic dataset, since their detection by shot gun mass spectrometry remains challenging. Some of the remaining candidates in our ESV dataset were upregulated on a post-transcriptional level during encystation [19] but are also expressed at significant levels in trophozoites [21]. Their analysis might bring to light additional proteins that are recruited to ESVs during encystation from primary localizations (e.g. the ER) in proliferating trophozoites. Altogether, our data provide strong indications that CWP1-3 are indeed the only highly abundant and completely stage-specifically expressed cyst wall components secreted via ESVs during encystation. Our results agree with previous data suggesting the formation of a minimal extracellular matrix in *G. lamblia*.

1.4. A proposal for cargo-driven ESV neogenesis

The lack of novel, ESV-specific protein candidates detected in this study suggests that ESV morphogenesis might be driven by a simple mechanism such as accumulation of cargo rather than by organelle-specific machinery, similar to the formation of dense-core secretory granules (DCSG) in eukaryotic cells.

DCSG comprise a key component in regulated exocytosis and are found in a wide variety of specialized eukaryotic cells. They are investigated in greatest detail in mammalian endocrine or exocrine cells where they act as intermediate storage compartments containing densely packed hormones or proteases and, in response to specific stimuli, fuse with the plasma membrane to release their contents [22, 23]. DCSG share, by definition, a core of electron dense material that appears dark in electron micrographs and granule maturation is often associated with proteolytic processing of cargo, for example cleavage of pro-insulin to insulin [23, 24].

Initially, two models have been proposed for the formation of DCSGs: The sorting-by-entry model, which is based on specific cargo selection by a membrane receptor [25-27], and the sorting-by-retention model, which describes the formation of immature granules containing both, granule-specific and unspecific cargo, while unspecific cargo is progressively removed from the granule via vesicle mediated trafficking [27].

The investigation of DCSG in organisms other than mammals has led to a steadily increasing number of secretory granules that do not conform to any of the two models, and there has been a lot of debate on where in the cell the granule cargo is actually sorted [23]. The diversity of secretory granules seems to be much more complex than originally thought. Nevertheless, protein sorting into most DCSG investigated to date appears to be associated with three features (figure 10) [23]: There are (i) membrane traversing or membrane-associated tethers, which in some cases recruit coat proteins on their cytoplasmic side, (ii) tether-associated cargo which is either captured directly or bound via linker proteins, and (iii) cargo components that have an inherent tendency to multimerize or aggregate and form condensed cores.

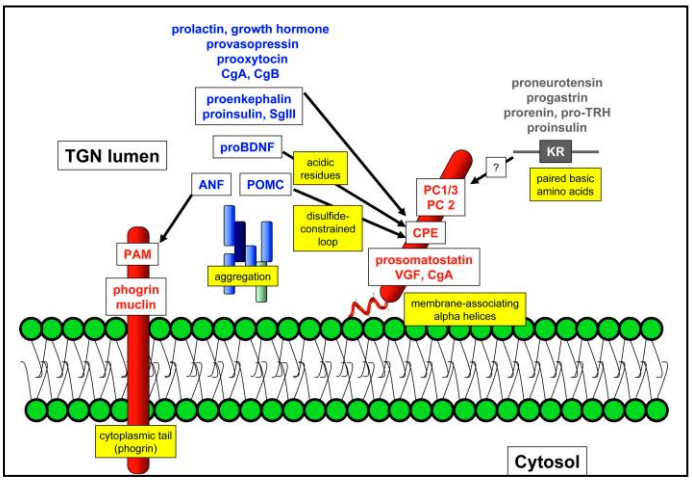


Figure 10: Model of protein sorting into DCSGs [23]. DCSG proteins can be divided into three groups: There are (i) thethers (red) that traverse or associate with the membrane, (ii) cargo proteins that associate with the membrane tethers, and (iii) cargo proteins that aggregate and form a dense core (blue). The core can be composed of several proteins. The yellow boxes indicate the different protein domains that have been involved in cargo sorting, and the white boxes indicate examples of cargo proteins.

CWP1-3 comprise the basic features required for ESV formation. **(i)** ESV formation at ERES has been shown to require the COPII-coat machinery [11, 12]. Because CWPs are soluble proteins, this suggests the presence of a membrane-spanning receptor which recruits COPII-coat proteins at the cytoplasmic side and selects CWPs in the ER lumen. A putative sorting signal for this receptor comprises the N-terminal region of CWP1, which was shown to be required to induce ESV formation in *Giardia* [28]. Alternatively, but

less likely, CWPs might act as membrane-tethers themselves by insertion of alpha-helical domains into the lipid bilayer or interact with modified membrane lipids, respectively. A dual role of proteins acting as both membrane tether and granule cargo is frequently observed [29-31] and would, in the case of CWPs, explain why no membrane receptor has been identified to date. **(ii)** CWP1 and the long processed domain of CWP2 are ESV cargo proteins that aggregate but are completely excluded from the dense core. These proteins build covalent complexes via rapid disulfide-crosslinking and isopeptide bonds [28, 32-34], but remain in a fluid phase during ESV maturation and formation of a dense core [35]. **(iii)** CWP3 has an inherent tendency to aggregate and to form a condensed core in maturing organelles together with the small cleaved domain of CWP2 [35]. Consequently, CWPs share basic hallmarks with most DCSG cargo from ciliates or animal cells [23].

The current data suggest the requirement of only simple (highly reduced) machinery driving ESV morphogenesis and maturation during the first 13 hours of encystation. According to the most basic scenario, CWPs induce formation of COPII transport intermediates by their inherent property to aggregate. In line with this, several microscopy studies suggest but could not definitively prove that ESVs are formed by progressive aggregation of electron-dense material in membrane-bounded compartments until large organelles were formed [13, 32, 36]. More recent studies show nucleation of ESVs in close proximity to ER exit sites [12].

The basic requirements for ESV morphogenesis by simple cargo aggregation are fulfilled; CWP3 was shown to form a condensed core [35], whereas CWP1 and CWP2 are known to undergo fast cross-linking via disulfide and isopeptide bonds [28, 32-34]. That inherent properties to aggregate might play a major role in ESV formation was suggested by data from several independent studies. Reiner and colleagues showed that the presence of dithiotreitol (DTT), which reduces disulfide bonds, leads to abandonment of ESV formation in trophozoites triggered for encystation and even disappearance of ESVs in already encysting trophozoites [33]. The reducing effect was reversible, since ESVs reappeared within minutes after removal of DTT. A more recent study did not confirm the disappearance of ESVs upon DTT treatment but showed that cross-linking of CWPs directly correlates with the morphology of the organelles: the authors observed a change in ESV compartment morphology from spherical to elongated, which was almost completely reversed after DTT was washed out [2]. Further, Davids and colleagues demonstrated the reduction of ESVs formation upon treatment with cystamine, an inhibitor of isopeptide-forming transglutaminases [34].

These data demonstrate that cross-linking of CWPs indeed is a key step in ESV morphogenesis and probably maintenance, and eventually is even sufficient to drive the entire process. In support of this scenario, co-expression of CWP1 and CWP2 in human embryonic kidney cells was shown to be sufficient to induce the formation membrane-enclosed vesicles which traffic the proteins to the cell surface (Abdul-Wahid 2004). This phenomenon is frequently observed upon ectopic over-expression of DCSG

cargo proteins, such as pro-vasopressin, chromogranin A, or von Willebrand factor in non-granule forming cells and strongly suggests that the cargo proteins themselves induce the formation of their proper carriers [37-39].

1.5. Where does the cyst wall sugar component join the game?

Considering its environmental resistance, the *Giardia* cyst wall is a biopolymer with a surprisingly simple composition consisting of three proteins (CWP1-3) and a β(1-3)-GalNAc homopolymer [40, 41]. The sugar component constitutes about 60% of the cyst wall mass and is synthesized from UDP-GalNAc monomers by an as yet unidentified cyst wall synthase [42]. While the CWPs are trafficked via ESV organelles to the surface of the cell, it is still unclear where and when the sugar component is synthesized, how it reaches the cell surface and how it is finally incorporated into the cyst wall.

It was proposed that UDP-GalNAc is synthesized from glucose in the cytosol by a series of enzymes that are upregulated during encystation (figure 11) [16]. The final step in synthesis of the monomer involves an UDP-GlcNAc-4'-epimerase (GALE) that converts a precursor, UDP-GlcNAc, to UDP-GalNAc [43].

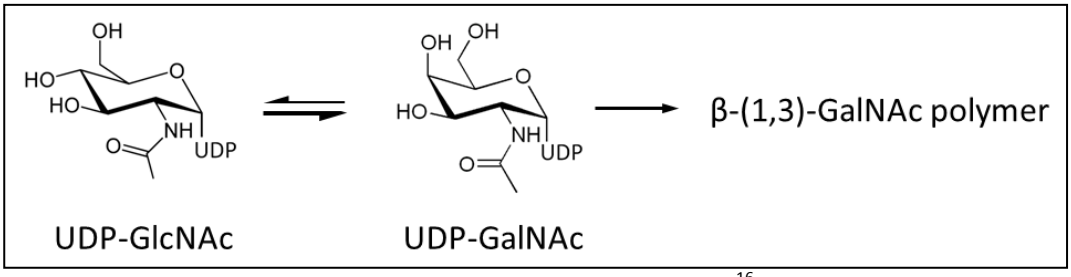


Figure 11: Synthesis of the *Giardia* cyst wall sugar homopolymer [16]. The precursor UDP-GlcNAc (left) is synthesized from endogenous glucose by a series of cytoplasmic enzymes. The precursor is converted into the cyst wall sugar monomer UDP-GalNAc (middle) by an as yet unidentified UDP-GlcNAc-4'-epimerase (GALE). The cyst wall sugar monomer is polymerized to the cyst wall sugar β-(1,3)-GalNAc polymer (right) by an as yet unidentified enzyme termed cyst wall synthase [42]. *UDP:* Uridine diphosphate; *GlcNAc:* N-acetyl-glucosamine; *GalNAc:* N-acetyl-galactosamine.

Although experimental data showed that the enzyme converts UDP-GlcNAc into UDP-GalNAc, the theory has a critical weak point because kinetic analysis revealed that the reverse reaction is clearly favored [44]. This raises significant doubts on the productive synthesis of the cyst wall monomer UDP-GalNAc by the proposed enzyme. Moreover, the synthesis of the cyst wall sugar in the cytosol would imply a transporter either in any organellar membrane (ER, ESVs, PVs) or the plasma membrane so that the sugar component reaches the cell surface – but an UDP-GalNAc transporter has not been identified in *Giardia* [45], and neither has a β(1-3)-GalNAc polymer transporter.

The lack of a lectin that specifically binds the unique β(1-3)-GalNAc polymer makes the investigation of the cyst wall sugar component challenging. A recent study demonstrated that CWP1 binds the GalNAc polymer of the *Giardia* cyst wall and used recombinant CWP1 to detect the sugar polymer [46]. Based on

these data, the authors suggested that the GalNAc polymer is made in vesicles distinct from ESVs and secreted early to the surface of encysting *Giardia* [46]. Although the idea of using CWP1 as a probe for detection of the sugar polymer is clever, the lack of control experiments and the presence of endogenous CWP1 in addition to epitope-tagged markers makes the interpretation of these data problematic.

Another study used the *Dolichos biflorus* agglutinin (DBA) to detect the GalNAc polymer [47]. Although the seed lectin has a much higher specificity for α-linked GalNAc, it can be generally considered as GalNAc-specific [48]. Although the data was of limited quality and not conclusive, Midlej and colleagues suggested that vesicles distinct from ESVs mediate transport of carbohydrate and perhaps also the GalNAc component to the surface of encysting cells [47].

In our study, two proteins in the ESV dataset attracted our attention since they might be involved in cyst wall sugar synthesis and transport: *Giardia*'s only nucleotide sugar transporter [49] which specifically transports cytosolic UDP-GlcNAc into the ER [45], and a putative UDP-GlcNAc-4'-epimerase which might convert UDP-GlcNAc into UDP-GalNAc (Wampfler and Hehl, under review). We detected the epimerase in the ER, whereas the transporter localized mainly to the perinuclear ER as well as to ESVs.

While there is strong support for import of UDP-GlcNAc into the ER for N-glycosylation, the presence of the epimerase producing UDP-GalNAc in the ER is intriguing; N-glycosylation in *Giardia* is restricted to addition of $GlcNAc_{1-2}$ to asparagine [50], hence a role of the epimerase in glycosylation is out of the question. One interpretation of our data is the involvement of the epimerase in UDP-GalNAc production for later cyst wall synthesis. This entails that the sugar component is either trafficked via separate vesicles to the surface of the cell, as suggested earlier [47], or reaches the surface via ESVs together with the CWPs.

1.6. Dense core secretory granules and the formation of extracellular matrices in protozoa

Dense-core secretory granules (DCSG) are found in many distantly related species and play important roles in diverse processes. The best characterized granules are those from animal endocrine or neuroendocrine cells, such as insulin-storage granules in pancreatic β-cells, or granules storing chromogranin A in chromaffin cells of the adrenal glands [23]. Besides DCSG of metazoa, the best investigated are those from ciliates, particularly the ones found in *Tetrahymena termophila* and *Paramecium tetraurelia* [53, 54]. Ciliate DCSG are often called toxicysts and release molecules that are used for hunting, as for example in catching and killing prey, or as defense mechanism against predators [55]. The social amoeba *Dictyostelium discoideum* secretes material via DCSG to form a resistant spore wall in order to survive extreme environmental conditions [56-58]. Even the phylogenetically most basic eukaryotic lineages possess DCSG. Similar to *Giardia*, *Entamoeba* [59-61] as well as apicomplexans such as *Eimeria* and *Toxoplasma* in their sexual stage [62] secrete components via DCSG to build environmentally resistant cysts or oocysts, respectively, which play a key role in surviving outside of the host. Invasive stages of *Eimeria*

and *Toxoplasma* synthesize three different secretory granules termed micronemes, rhoptries, and dense granules, and secrete their contents sequentially during and after host cell invasion [63-66].

Three examples of secretory granule biogenesis and the formation of protective extracellular matrices in protozoa are discussed in more detail below: the oocyst wall formation in *E. maxima*, encapsulation of *T. termophila*, and encystation of *E. invadens*.

Oocyst wall formation in Eimeria maxima

Similar to *Giardia*, coccidian parasites such as *Eimeria* and *Toxoplasma* rely on the synthesis of a robust oocyst wall for their fecal-oral transmission from host to host. *E. maxima* stores the material for oocyst wall formation in three specific organelles which are developed during the parasite's sexual, macrogamete stage: the veil forming bodies (VFB) and the wall forming bodies type 1 (WFB1) and type 2 (WFB2) [67, 68]. The oocyst wall is formed upon sequential release of the contents of these organelles and was described in detail by Ferguson et al. [69].

VFB secrete their contents already in an early macrogamete stage, giving rise to a loose veil around the oocyst. WFB1s and WFB2s are synthesized only during macrogamete maturation (figure 12). WFB1s are large spherical compartments and contain electron-dense material [62]. They emerge during macrogamete maturation and locate to the peripheral cytoplasm, but the origin of the organelles or how their contents are trafficked, respectively, was never determined [62]. WFB2s are less electron-dense and develop as dilations of the rough ER. In early stages of macrogamete maturation they have amorphous content which later develops into a characteristic, electron-dense and doughnut-shaped structure [62]. The WFB2s remain enclosed in the rough ER during macrogamete maturation. In the mature macrogamete both, the WFB1s and WFB2s, locate to the periphery of the cell (figure 12, g). Whether the WFB2s emerge directly from the rough ER or if their content is transported via the Golgi to the peripherally located organelles was never determined [69]. The oocyst wall material is secreted sequentially when the mature macrogamete is fertilized by a microgamete and develops into a zygote. Fusion of the cytoplasmic WFB1s with the cell surface, and apparent polymerization of the material, results in the formation of an outer, electron-dense oocyst wall. At the same time, the WFB2s lose their doughnut-shaped structure and disappear simultaneously with the emergence of an inner oocyst wall (figure 12, j and l).

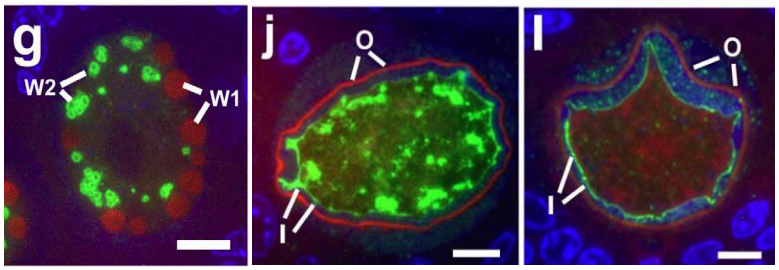


Figure 12: Different stages of oocyst wall formation in *Eimeria maxima* [69]. **(g)** Mature macrogamete with stained wall forming bodies (WFBs) containing the material for later oocyst wall formation. The WFB1s (red) are stained with Evans Blue, whereas the WFB2s (green) are stained with an antibody against the WFB2 cargo protein Gam56. Gam56 shows the typical doughnut-formed distribution in WFB2s observed in the mature macrogamete stage. Both WFBs locate to the cell periphery at this time point. **(j, l)** Fully formed oocysts walls. The outer layer is stained with Evans Blue (red), whereas the inner layer is stained by either anti-Gam56 (j, green) or anti-Gam230 (l, green) antibodies. Some residual cytoplasmic staining by anti-Gam56 (j) is visible.

The *Eimeria* oocyst wall formation has striking parallels with the encystation process in *Giardia*. Both parasites store their wall material in granules and sequentially secrete it to the surface to build first an outer and subsequently an inner layer of the wall. Another similarity is the formation of biophysically distinct compartments within granules, visible by the doughnut-shaped distribution of CWP1 and CWP2 surrounding a condensed core in ESVs and the Gam56 and Gam82 proteins in the *Eimeria* WFB2s. Analogous to *Giardia* CWP2, the *Eimeria* WFB cargo proteins Gam56 and Gam82 are proteolytically processed before they are integrated into the oocyst wall [62]. However, the parasites show an interesting difference in secretion of their wall components: While *Eimeria* mediates sequential secretion by spatial separation of the material in different granule types (VFB, WFB1, WFB2), *Giardia* solves this issue by the formation of two biophysically distinct fractions in a single organelle type (ESVs).

Encapsulation of Tetrahymena termophila

The ciliate *Tetrahymena termophila* secretes material in a similar way in order to build a sturdy capsule around the cell. Since the capsule does not enclose the entire cell an environmental protective function is unlikely, and it was therefore suggested that the capsule is used for nutrient trapping or defense against predators [55]. The capsule material is stored in more than 100 dense-core granules termed mucocysts (figure 13, A-C). Detailed microscopy studies initially proposed the emergence of the granules from ER-derived membranes [70]. In the light of current information, it is generally assumed, however, that mucocysts are formed at the TGN [71], probably because animal DCSG, as the most characterized ones to date, are formed at the TGN[23, 72].

The six most abundant *T. termophila* mucocyst proteins identified belong to a family named GRL (GRanule Lattice) proteins [71, 73, 74]. In addition to the GRL proteins a second, poorly understood class of 13 luminal proteins was identified, including GRTp1 (GRanule Tip protein 1) which preferentially localizes near the granule tip [75]. Similar to CWPs in *Giardia*, the *T. termophila* GRL proteins are initially assembled in the ER where they form hetero-oligomers. Deletion of each individual GRL protein results in the ER

retention of the remaining GRL proteins [74]. This indicates that, similar to *Giardia*, aggregation of GRL cargo proteins seems to play a key role during their export from the ER. In further analogy with ESVs, mucocyst maturation is characterized by proteolytic processing and selective condensation of the GRL proteins into a crystalline core [76, 77]. During the maturation process, mucocysts are distributed in the cytosol (figure 13A). In later stages, the mature granules are docked to the plasma membrane (figure 13, B and C).

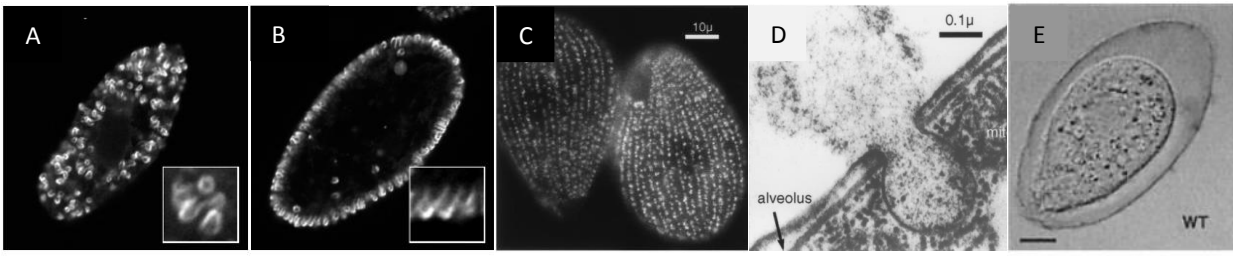


Figure 13: Encapsulation of *Tetrahymena termophila.* **A, B)** For encapsulation, *T. termophila* synthesizes more than 100 granules, termed mucocysts, that contain material for the capsule. Mucocysts are visualized by indirect immunofluorescence of the granule cargo protein GRT1 (granule tip protein 1) which is concentrated at the tip of each granule. The six GRT proteins which are part of the condensed granule core are not visualized [53]. Immature granules are distributed in the cytosol of the ciliate (A), whereas mature granules are docked to the plasma membrane (B). **C)** Immunofluorescence staining of the mucocyst cargo protein p80 visualizes mucocyst granules arrayed as meridians covering the entire surface of the organism apart from the mouth structure [78]. **D)** Exocytosis of a mucocyst shows fusion of the mucocyst membranes with the plasma membrane and expansion and expulsion of the mucocyst contents [78]. **E)** Light micrograph of an encapsulated *T. termophila* cell [73].

Mucocyst fusion with the membrane can be triggered by a rise in cytosolic calcium and results in synchronous cargo release within seconds (figure 13D) [78]. During exocytosis, the crystalline structured GRL proteins in the dense core undergo a roughly 7-fold expansion [76], resulting in the rapid encapsulation of the parasite (Figure 13E). The expansion of the so called “spring-loaded” GRL proteins is controlled through their binding to calcium [76]. The calcium sensitivity of the GRL1 protein remarkably increases after proteolytic processing, and the processed form of GRL1 has been shown to undergo a conformational change upon calcium binding [79]. In contrast to the GRL proteins, the GRT proteins seem not required for mucocyst cargo assembly or secretion of the capsule material, thus a post-exocytic function was suggested for this second family of cargo proteins [79].

T. termophila encapsulation shows many parallels to *Giardia* encystation. However, a difference lies in the de-condensation of the dense core components. Whereas the *Tetrahymena* GRL proteins are “spring-loaded” and undergo fast de-condensation within seconds after exocytosis, secretion of the *Giardia* CWP3 core takes several hours, suggesting a slow de-condensation process prior to exocytosis [35].

Encystation in Entamoeba invadens

Entamoeba histolytica is the cause for amebic dysentery and liver abscesses in humans. Analog to *Giardia*, host transmission of *E. histolytica* occurs via the fecal-oral route of the infectious stage, the cyst. Since in vitro triggering of encystation in *E. histolytica* is highly inefficient, its reptile-infecting relative *E. invadens* has become the reference model for *Entamoeba* encystation [80].

The E. invadens cyst wall is comprised of a β-(1,4)-GlcNAc homopolymer (chitin) [81] and a small set of chitin-binding proteins termed Jacobs, Jessies and chitinases [59, 60]. The encystation process involves sequential secretion of the cyst wall sugar and protein components via separate vesicles. The encystation process was visualized and described in detail by Chatterjee et al., and based also on additional data from important previous works [59, 60], the authors suggested a "wattle and daub" model for cyst wall formation (figure 14).

Early during encystation, two types of vesicles of unknown origin containing Jacob lectins and chitin arise in the cytoplasm. Gradual secretion of their content to the cell surface results in the formation of an as yet immature cyst wall which is still permeable to small molecules (figure 14, wattle stage). Due to the cyst wall permeability at this stage, Chatterjee et al. suggested that chitin and the chitin-binding Jacob proteins build the scaffold or wattle of the cyst wall. Jacobs, which are glycoproteins themselves, are linked to the plasma membrane by constitutively expressed surface lectins [59].

In a later stage, an additional set of vesicles containing the Jessie lectins is synthesized. Secretion of Jessie lectins and integration into the wall is accompanied by the reduced permeability of the cyst wall to small molecules, suggesting that Jessie lectins play a role in sealing the cyst wall (figure 14, daub stage). Chatterjee et al. assumed that Jessies are linked to the cyst wall via their chitin binding domain and seal the membrane by their tendency to self-aggregate [61]. However, the authors did not rule out other mechanisms, such as disulfide-crosslinking, which might affect cyst wall permeability. The role of chitinases, which are secreted early during encystation, is largely unknown.

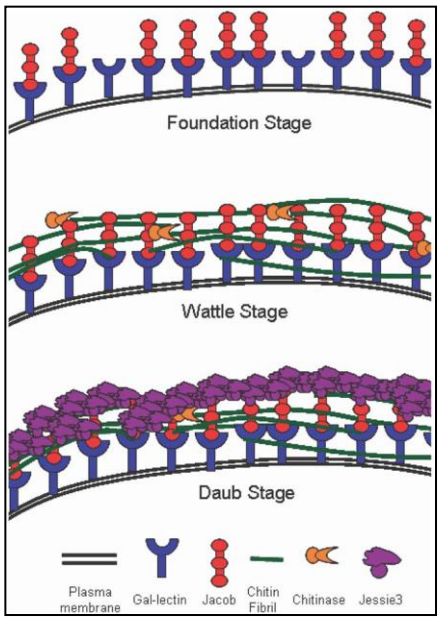


Figure 14: The wattle and daub model for encystation in *Entamoeba invadens* [46]. The model predicts accumulation of the *E. invadens* cyst wall in three stages. 1) Foundation stage: the Jacob lectins (red), which are themselves glycoproteins, are secreted to the surface and linked to the plasma membrane by constitutively expressed integral membrane lectins (blue). 2) Wattle stage: chitin (green) and chitinases (orange) are deposited on the surface. Jacob lectins cross-link the chitin fibrils via several chitin-binding domains. 3) Daub stage: the cyst wall gets impermeable by the integration of Jessie lectins (violet). Jessie lectins are linked to the cyst wall by their chitin-binding domain. Sealing probably occurs due to their tendency to self-aggregate.

The encystation processes in *Entamoeba* and *Giardia* have several striking parallels. The cyst wall of both parasites is made up of a carbohydrate homopolymer, which is the β-(1,4)-GlcNAc polymer chitin in *Entamoeba* and a β-(1,3)-GalNAc polymer uniquely found in *Giardia* [40, 41, 81]. The limited data available generated without a specific marker for the cyst wall glycan suggest that the cyst wall sugar component in *Giardia* is secreted by vesicles distinct from ESVs [47], which would be analogous to the secretion of chitin in *Entamoeba*. Further similarities are found for the *Entamoeba* and *Giardia* CWPs themselves. Some or all CWPs have lectin activity and bind to the cyst wall sugar component and have tendency to self-aggregate [35, 46, 61]. In addition, some of the proteins are proteolytically processed during encystation [35, 60, 82].

The existence of DCSG in many different species reflects the need for rapid and regulated secretion under certain conditions for survival. Animal DCSG share many fundamental hallmarks with DCSG from ciliates or other protozoa, such as granule maturation by proteolytic processing, formation of a dense core, and calcium-triggered, regulated exocytosis mediated by SNARE proteins and Rab GTPases [22].
Their widespread distribution among species including the most ancient eukaryotes and their common hallmarks suggests that DCSG evolved from an early eukaryotic ancestor. However, molecular studies on DCSG of distantly related species strongly argue that DCSG in different lineages were invented several times independently and are the product of convergent evolution [22]. Elde and colleagues demonstrated that key DCSG proteins, such as luminal cargo proteins, processing proteases, and integral membrane proteins involved in granule docking and exocytic fusion, are clearly related in the ciliates *Paramecium* and *Tetrahymena*, but do not show homology to their animal counterparts. The same is true for DCSG-associated Rab and SNARE proteins which mediate granule docking and fusion of ciliate or animal DCSG [22]. The missing relatedness of granule-associated Rab GTPases and SNARE proteins between animals and ciliates is particularly informative because these proteins are usually highly conserved between even distantly related species. Consistent with this, Rab and SNARE proteins associated with the Golgi of ciliates have Golgi-associated orthologs in animals [22].
Convergent evolution of DCSG in diverse species may be the result of selective pressure on the regulated secretory pathway and adaptation of the organisms to different environmental conditions. Consistent with this idea, DCSG of different species are involved in a wide variety of processes and secrete diverse cargo, allowing individual species to perfectly adapt to their specific biological niches.

1.7. Three scenarios for cargo-driven ESV neogenesis

During the past 20 years, considerable effort was undertaken to discover a defining, ESV-specific organelle marker. Despite substantial effort on all levels, including extensive genome searches [14, 52], transcriptome studies [4, 18], and proteomic analysis of ESV organelles and even the entire *Giardia* proteome at different life cycle stages [2, 19, 83], no defining organelle-marker was identified. Based on these data and our unsuccessful attempt at detecting unique ESV-specific proteins (Wampfler and Hehl,

under review), we suggest three non-mutually exclusive scenarios for morphogenesis and maturation of ESV organelles with CWPs as the main driving factors. The models are schematically illustrated in figure 15, and their main difference lies in the initial cargo sorting step:

Model 1: An unknown integral membrane receptor captures CWP1 and recruits the COPII coat to the ER

Model 2: An unknown integral membrane receptor captures the nascent cyst wall UDP-GalNAc homopolymer and recruits the COPII coat to the ER

Model 3: CWPs directly interact with the ER membrane, either by insertion of an alpha-helical domain into the lipid bilayer or via interaction with modified lipids.

Model 1

In model 1 (figure 15, top), we suggest the participation of an integral membrane receptor that captures ESV cargo. The probably strongest indication for the presence of such a receptor for CWPs is the fact that soluble, secreted proteins in most cases depend on cargo receptors for recruitment of cytoplasmic membrane coat complexes for export from the ER.

The dependence of ESV formation on COPII coats was mainly experimentally demonstrated in two studies. Stefanic and colleagues created a *G. lamblia* cell line expressing a dominant negative mutant variant of the COPII coat recruitment GTPase Sar1 [11]. These transgenic cells formed only few ESVs with aberrant morphology, if any, strongly suggesting that export of CWM from the ER and thus ESV formation depends on the COPII coat machinery. Later, Faso and colleagues nicely visualized the neogenesis of ESVs at ER exit sites in *Giardia* [12]. Whether a predicted integral membrane receptor captures CWPs was not determined. However, the finding that the N-terminal domain of CWP1 is required for ER export and protein targeting to ESVs implies a putative sorting domain within CWP1 [28].

Covalent and non-covalent intermolecular interactions between CWPs provide unique opportunities for co-transport and sorting of CWPs into ESVs [28, 32-34]. CWPs form intermolecular complexes which are captured and retained by a specific membrane receptor, and newly synthesized CWPs are progressively cross-linked and retained in growing ESVs in a “sorting by retention” mechanism. CWP3 and the short proteolytic fragment of CWP2 form a solid core based on their inherent property to aggregate, and perhaps with the help of additional as yet unknown factors [35].

Model 2

Analogous to model 1, model 2 (figure 15, middle) involves an integral membrane receptor recruiting the COPII coat machinery on the cytoplasmic side. Instead of capturing CWPs directly, the model posits that the cargo receptor is a lectin that selectively binds a nascent cyst wall GalNAc homopolymer.

Our organelle proteome study suggests that the UDP-GalNAc monomer might be converted from UDP-GlcNAc in the ER (Wampfler and Hehl, under review). Thus, formation of the GalNAc homopolymer could be initiated in the ER, where it is captured by a receptor with specific lectin activity. The carbohydrate in turn captures soluble CWPs [46], packaging all major components for export.

The *Giardia* cyst wall synthase is deemed the responsible enzyme for polymerization of UDP-GalNAc monomers into the β(1-3)-GalNAc polymer [42]. Although the enzyme's activity has been detected in a particle fraction isolated by isodensity fractionation from encysting *Giardia*, its identity as well as its localization within compartments was never determined. Of interest, the authors considered a localization of the enzyme in ESVs, since activity of the synthase was detected in fractions that also labeled positively with monoclonal antibodies against CWP1. However, ESV-mediated export of the synthase could also mean that the enzyme gets a ride to the cell surface and does not automatically imply its activity at this stage.

Our model 2 is in agreement with the export of the cyst wall carbohydrate from the ER which was suggested by Midley and colleagues [47], but contrasts in the sense that secretion of the carbohydrate occurs via ESVs and not in independent vesicles [47].

Model 3

Model 3 (figure 15, bottom) posits that CWPs directly interact with the membrane by the insertion of an amphiphatic alpha-helix into the lipid bilayer. Interactions of this kind are observed frequently for DCSG cargo proteins [23]. Another possibility comprises the interaction of CWPs with specific membrane lipid domains [72].

A direct interaction of CWPs with the membrane would explain why no cargo receptor has been identified to date. However, the lack of a receptor in model 3 raises the question how the COPII coat is recruited to the budding transport intermediate. Only recently, several different mechanisms for export of bulk cargo from the ER were unraveled [84-87]. As an example, Siddiqi et al. identified two unusual budding complexes transporting large lipoprotein particles with diameters up to 250 nm from the ER to the Golgi in rat hepatocytes and intestinal cells [84-86]. Although the complexes involve one or more COPII proteins, they contain protein markers distinct from the canonical COPII coat machinery. Even more interesting, Thor and colleagues showed that rapid, COPII-dependent bulk flow from the ER in ovary hamster cells is induced even in the absence of specific export signals [88].

Alternatively, we have to consider the involvement of lipid rafts. These have been shown to facilitate coat assembly at the TGN when secretory granules were formed [72]. Lastly, we should consider that COPII

coat recruitment to budding ESVs in *Giardia* might follow an unconventional manner, and does not absolutely require the involvement of a receptor.

Physical links between ESVs and the ER

Based on data from our study, ESVs maintain a strong and persistent contact with the ER. We predict that the two organelles remain connected for a long time, probably even until cargo processing and partitioning is finished at ~12-14 hours p.i., before sorting occurs [35]. During this period, distinct ER-resident proteins, such as Hsp70-BiP, a high cysteine non-variant cyst protein (HCNCp), several EGF-like cyst protein family members (EGFCP 1-6), and the novel candidates identified in our study, appear to have access to ESVs (Wampfler and Hehl, under review, [2, 20, 89]). Because the majority of these proteins are not secreted to the surface, they must be cycled back to the ER at a certain time. This might occur via low density, COPI coated vesicles. Such a "granule clean up" is frequently observed when DCSG bud off from the TGN in mammalian cells, giving rise to the originally postulated "sorting by retention" model [27]. In analogy, export of the *Giardia* chaperone Hsp70/BiP, which has a KDEL retention signal, and its subsequent retrograde transport to the ER most likely in a ERD2- and COPI-dependent manner was experimentally demonstrated [2]. Alternatively, removal of the proteins from ESVs could also occur via tubular membrane links between ESVs and the ER [11]. These tubular connections may also be involved in exchanging fluid phase cargo between ESVs and were suggested to synchronize maturation of the organelles.

Cargo maturation

During ESV maturation, CWP2 is proteolytically processed (starting at ~8h p.i.) and partitioned in two fractions; whereas the long N-terminal part remains in a fluid state together with CWP1, the small C-terminal fragment of the protein undergoes condensation with CWP3 and localizes to the dense core of ESVs [35]. There is an ongoing debate about the protease responsible for CWP2 processing, with three postulated candidates [90-93].

A lasting connection between ESVs and the ER until cargo processing and granule maturation is completed suggests that the protease, which cleaves CWP2, localizes to the ER or to budding ESVs. This has only been shown for one of the suggested candidates, a subtilisin-like proprotein convertase (SPC), which localizes mainly to the perinuclear ER in trophozoites and to ESVs in encysting cells [93].

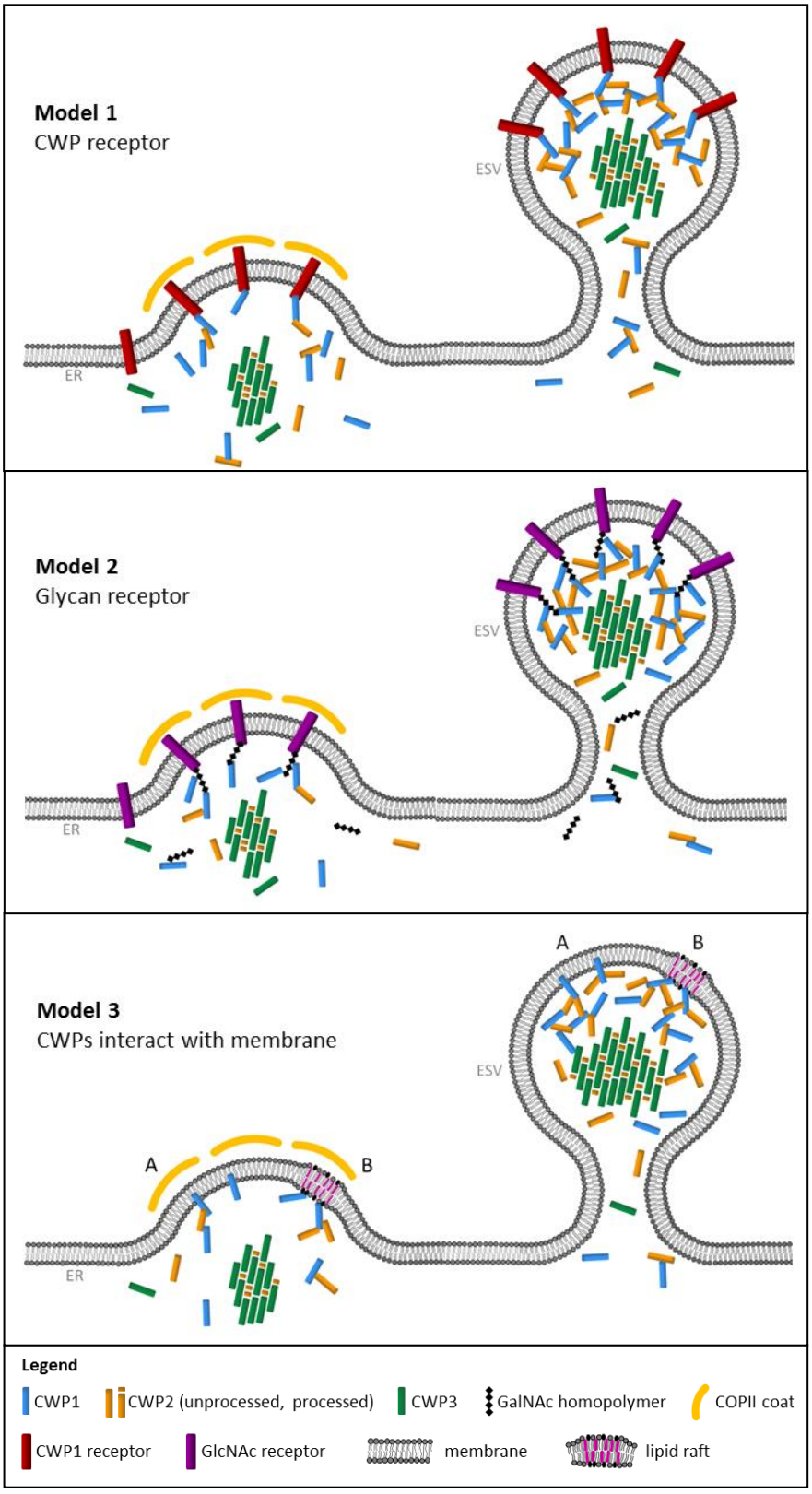


Figure 15: Three proposed models for ESV morphogenesis in *G. lamblia*. **Model 1:** An integral membrane receptor captures CWPs (example: CWP1) and recruits the COPII coat at the cytoplasmic side. The coat mediates membrane bending at an early stage and is lost when ESVs mature. CWP2 enters the granule by disulfide crosslinking with CWP1, and CWP3 undergoes condensation with the small proteolytic fragment of CWP2.

Model 2: An integral membrane receptor captures the nascent GalNAc carbohydrate polymer and recruits the COPII coat at the cytoplasmic side. After initial membrane bending for granule formation the coat is disassembled. The GalNAc polymers recruit CWPs by binding to their lectin domains. CWP3 aggregates together with the small proteolytic fragment of CWP2 in a condensed core.

Model 3: CWPs (example: CWP1) directly interact with the membrane, either by insertion of an alpha-helical domain (**3A**) or by interaction with lipid-raft like domains (**3B**). CWP2 enters the granule by disulfide cross-linking with CWP1, and CWP3 undergoes condensation with the small proteolytic fragment of CWP2. The COPII coat is recruited to early ESVs by alternative mechanisms.

During our studies we got strong indication that formation of ESVs is primarily driven by cargo proteins themselves and does not involve organelle-specific factors. Based on these data we suggested three conceivable models for ESV neogenesis and maturation with CWP1-3 as key factors.

Despite considerable progress in recent years and the availability of several working models, the molecular underpinning of ESV neogenesis and regulation are still incompletely understood: A number of important questions await experimental investigation: How is cargo sorting into ESVs mediated? How is selective condensation achieved? At which stage are ESVs independent from the ER? Where and how is the glycan component integrated into the cyst wall?

A this point it becomes clear that addressing these questions requires new approaches, for example knock-out of the key players CWP1-3. However, the parasite's tetraploid organization [94, 95] and the availability of only two suitable resistance markers for selection of transgenic cells [96-99] makes genetic manipulations which are routine in other systems, such as gene knock outs, unfeasible. Alternative approaches to reduce protein levels such as expression of antisense sequences [11, 100-103], long double-stranded RNA [104, 105], or transfection with translation-blocking morpholinos [106] resulted in variable and inconsistent knock down levels [11, 105-107]. In order to overcome some of these limitations we established the *E. coli* phage P1-derived Cre/loxP system in *G. lamblia* (Wampfler and Hehl, manuscript in preparation). In addition, we provide a detailed protocol how this newly established system can be used to realize gene knock outs in the tetraploid parasite.

1.8. The Cre/loxP system in *G. lamblia*

The Cre-recombinase is an *E. coli* phage P1-encoded enzyme that mediates recombination of short DNA sequences termed loxP sites, helping the phage to maintain its circular DNA as a single unit copy in the host cell and thus ensuring proper segregation to daughter cells [108, 109]. Since its characterization 30 years ago [108, 110], this simple and efficient two-component system has become a powerful tool in genome research and is applied to tightly control genome manipulations in diverse organisms, including protozoan parasites such as *T. gondii*, *T. brucei*, and *P. falciparum* [111-113].

Investigation of gene and protein function in *Giardia* is currently restricted to (over)expression of (modified) target genes or the reduction of protein levels by interfering with mRNA synthesis, stability, or accessibility to the translation machinery [11, 104, 106, 114]. However, knock-down approaches reported so far, including the transfection with morpholinos and hammer head ribozymes as well as the expression of antisense sequences or double-stranded RNA, commonly result in inconsistent levels of protein depletion [11, 105-107, 114].

To address the challenge of manipulating *Giardia*'s tetraploid genome in a more effective manner, we implemented the Cre/loxP system in this binuclear protozoan parasite (Wampfler and Hehl, manuscript in preparation).
In order to test if the Cre-recombinase is expressed and functional in *G. lamblia*, we designed a Cre/loxP-based reporter system consisting of a split GFP ORF interrupted by a floxed mcherry expression cassette (Wampfler and Hehl, manuscript in preparation). We showed that, upon activation of Cre, the floxed mcherry expression cassette is excised from the *Giardia* genome, resulting in reconstruction of a full length GFP ORF and expression of the GFP-reporter. We then used the new tool to excise a stably inserted, floxed puromycin resistance from the *Giardia* genome and demonstrated that the same selectable resistance marker can be recycled and used for a second round of transfection in the same cell line.

The major limitation of the Cre/loxP system in *Giardia* comprised the long maintenance of the Cre-encoding plasmid in the cultured cells even in the absence of positive selection pressure. The interpretation of our data suggests that once activated, the recombinase established equilibrium between excision and insertion of the floxed puromycin resistance cassette, surprisingly despite the presence of a tightly regulated inducible promoter controlling Cre-expression. The equilibrium, which is driven towards excision, stagnated only upon loss of the Cre-encoding episome in cultured cell populations after more than 4 weeks.

Our results demonstrated that conditional expression of the Cre-recombinase by the positioning of an inducible promoter sequence upstream of the Cre-coding region is not controlled tight enough. In theory, only 4 Cre molecules per cell are required for recombination of two loxP sites and hence an establishment of equilibrium between excision and insertion [115]. Of interest, Andenmatten and colleagues [116] developed an alternative strategy in *T. gondii* that permits regulation of the Cre-recombinase on a protein level (figure 16). The DiCre system is based on the conditional expression of the Cre-recombinase as two separate, N-terminal and C-terminal, subunits which are either fused to the FK506-binding protein (FKBP) or the rapamycin binding domain of mTOR (FRB). Addition of rapamycin, an anti-fungal antibiotic, causes the formation of a stable FKBP-rapamycin-FRB complex [117], resulting in dimerization of the two Cre-subunits and thus activation of the recombinase.

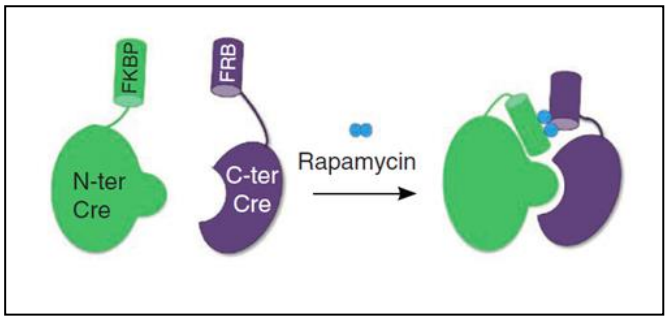


Figure 16: The DiCre system [116]. **Left:** The N-terminal subunit of Cre is fused to FKBP12 (green), and the C-terminal subunit of Cre is fused to FRB (violet). The addition of rapamycin leads to the formation of a FKBP-rapamycin-FRB ternary complex and dimerization of the Cre subunits, reconstituting its enzymatic activity. *FKBP:* FK506 binding protein, *FRB:* Rapamycin-binding domain of mTOR (mammalian target of rapamycin).

A crucial finding was that the Cre-recombinase tolerates certain nucleotide variations in its target sequences, but recombines only homotopic pairs of these variant sites [115] (figure 17). The combined application of several alternative target sites, such as lox511 [118], lox2272 [119], or m2[120], permits the realization of complex manipulations of the genome.

Target Site	Inverted repeat 1	Spacer	Inverted Repeat 2
*lox*P	ATAACTTCGTATA	ATGTATGC	TATACGAAGTTAT
*lox*511	ATAACTTCGTATA	ATGTATaC	TATACGAAGTTAT
*lox*5171	ATAACTTCGTATA	ATGTgTaC	TATACGAAGTTAT
*lox*2272	ATAACTTCGTATA	AaGTATcC	TATACGAAGTTAT
m2	ATAACTTCGTATA	AgaaAcca	TATACGAAGTTAT
*lox*71 (LE mutant)	taccgTTCGTATA	ATGTATGC	TATACGAAGTTAT
*lox*66 (RE mutant)	ATAACTTCGTATA	ATGTATGC	TATACGAAcggta

Figure 17: Alternative Cre targeting sites. The wildtype loxP site consists of an 8 bp asymmetrical core (spacer) flanked by two 13 bp inverted repeats [110]. Alternative targeting sites contain nucleotide variations in the spacer sequence (lox511, lox5171, lox2272, and m2) or in the inverted repeats (lox71, lox66) and are tolerated by the Cre-recombinase [115].

When used as inverted repeat variants, these alternative target sites might provide a straight-forward strategy to overcome the challenge of Cre-mediated equilibrium in *Giardia*. Inverted repeat-variants, termed LE (left element) and RE (right element) mutants, contain nucleotide variations in their left or right inverted repeat, respectively (figure 17, bottom). Cre-mediated recombination is effectively performed between LE and RE mutant sites, but results in the generation of two different target sites that cannot be recombined a second time (figure 18). Inverted repeat-variants were mainly used to generate stable DNA insertions [115].

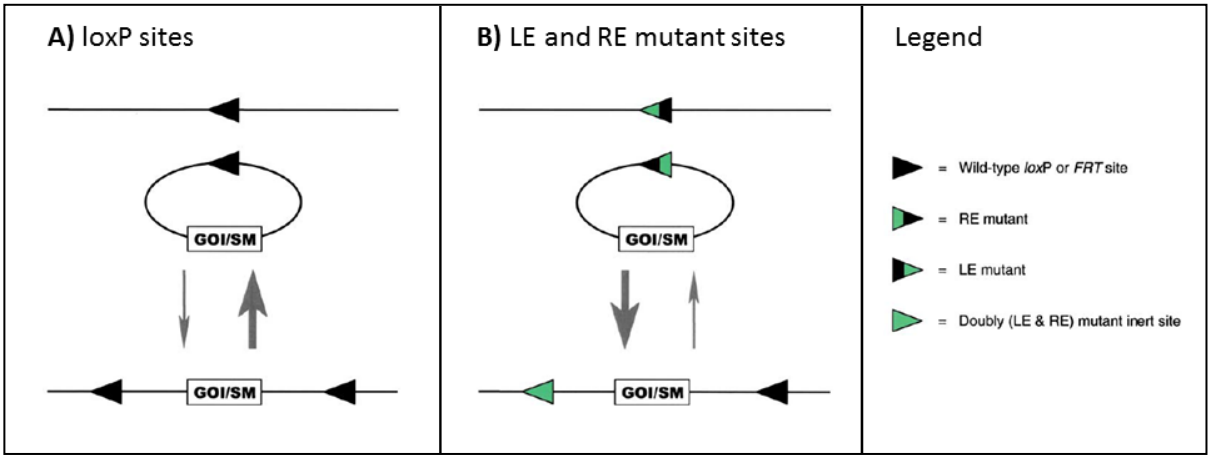


Figure 18: Use of loxP inverted repeat variants to achieve stable DNA insertion [115]. **A)** Cre-mediated recombination between two wildtype loxP target sites can be performed repeatedly, resulting in equilibrium between excision and insertion of the floxed DNA sequence. **B)** Inverted repeat variants containing single bp substitutions in either the left inverted repeat (left element, LE) or the right inverted repeat (right element, RE). Cre-mediated recombination of LE and RE mutant sites results in the generation of two different, inert target sites which are not recombined efficiently with each other. *GOI:* gene of interest; *SM:* selectable marker, *RE:* right element (right inverted repeat); *LE:* left element (left inverted repeat).

Application of inverted repeat-variants in *Giardia* by positioning LE and RE target sites on either side of a puromycin resistance cassette would permit fast and permanent removal of the resistance marker from the *Giardia* genome, hence considerably accelerate the system when applied for serial genome manipulations.

The Cre/loxP technique in *G. lamblia* provides novel means for more complex and versatile genetic manipulations, and it will facilitate future investigation of gene and protein function in this parasite. In conjunction with alternative Cre-recombinase target sites, this genetic tool has the potential to overcome some of the major barriers in *G. lamblia* research, such as the parasite's tetraploid genome and the limited repertoire of selectable markers, and brings us a step closer towards the first gene knockout in this evolutionary diverged protozoan.

2. Perspectives

During our studies we got strong indications that morphogenesis of ESVs is primarily driven by cargo proteins without involving organelle-specific factors. Based on these data we proposed three models for ESV neogenesis and maturation. However, many open questions remain, some of which can only be addressed by gene knock outs of the key players CWP1-3. With the implementation of the Cre/loxP system in *G. lamblia* we developed the required technology and toolbox to generate gene knock outs in this tetraploid parasite. Hereinafter we describe how a CWP1 knock out, as an example, could be realized in *G. lamblia*.

2.1. CWP1 gene knock out in *Giardia lamblia*

To perform a gene knock out in the tetraploid *Giardia* [94, 95], four rounds of homologous recombination are required to remove all gene copies. So far, the availability of only two resistance markers, puromycin and neomycin [96-99], did not allow this. Here we propose the repeated use of a floxed puromycin resistance cassette which is excised from the genomic insertion site by Cre-recombinase after each round of homologous recombination (Wampfler and Hehl, manuscript in preparation). A schematic description of the strategy is provided in figure 19. For a proof of concept we chose one of the CWPs because the protein is likely completely dispensable in trophozoites and no conditional knock out strategy is therefore required.

Specifically, we would design four plasmids, each containing a floxed puromycin resistance cassette flanked by different recombination sites (RS1 – RS4) that target the construct to the CWP1 gene locus. In a first round of recombination, one copy of CWP1 is exchanged by a floxed puromycin resistance cassette for selection of transgenic cells, while three copies of the CWP1 gene remain in the genome (figure 19, A1). The Cre-recombinase, under control of the inducible CWP1-promoter, is then transfected into the cells via an episomally maintained plasmid using neomycin selection. Transient activation of the recombinase by activating the CWP1 promoter controlling its transcription (Wampfler and Hehl, manuscript in preparation) results in Cre-mediated excision of the floxed puromycin resistance (figure 19, A3). The Cre-encoding plasmid is lost over time due to termination of antibiotic pressure by neomycin. To obtain clonal cell lines that successfully removed the plasmid, flow cytometry-based single cell sorting can be performed.

In the following rounds of homologous recombination, one has to ensure that the construct is targeted to a remaining intact copy of the CWP1 target gene and does not recombine with the “empty” gene locus where previous crossover reactions took place. This is achieved by using different recombination sites in each round, i.e. flanking upstream and downstream sequences in the target gene region but at different distances (figure 19, B and C). Each recombination reaction therefore removes not only the

target gene but also all sequences upstream and downstream of the target gene region within the recombination sites. Consequently, the number of possible recombination sites in each round decreases, “forcing” the plasmid to replace another intact CWP1 gene copy.

Each Cre-mediated excision of a floxed puromycin resistance leaves one loxP site behind in the genome. For each new recombination round, a plasmid with two additional loxP sites flanking the puromycin resistance is inserted into the cell. Since the Cre-recombinase does not differentiate between identical target sites, more than two loxP site would result in uncontrolled insertion, excision, or transformation events. In order to ensure the proper removal of floxed puromycin in each round, alternative loxP sites must be used. The Cre-recombinase recognizes these alternative target sites but mediates recombination specifically only between homotopic ones [115]. For example, one could use loxP and the alternative sites lox5171, lox2272 and m2 [118-120].

Cre-mediated excision of floxed DNA sequences is not a uni-directional reaction but instead an equilibrium between excision and insertion [115]. As long as Cre is present, a subpopulation of the cells will maintain the floxed puromycin resistance in the genome (Wampfler and Hehl, manuscript in preparation). Consequently, one has to wait after each knock out round until the episomally maintained plasmid encoding Cre has been lost. An alternative approach is the transfection of cells with the Cre-recombinase protein, analogous to the restriction enzyme mediated integration (REMI) technique [121]. The advantage is that proteins taken up by the cells are not replicated and are thus rapidly diluted when cells divide. However, in the absence of any selection pressure this advantage most probably goes to the expense of decreased transfection rate and recombination efficiency. Alternatively, the Cre recombinase activity itself could be regulated, using the recently developed DiCre system in *T. gondii* [116].

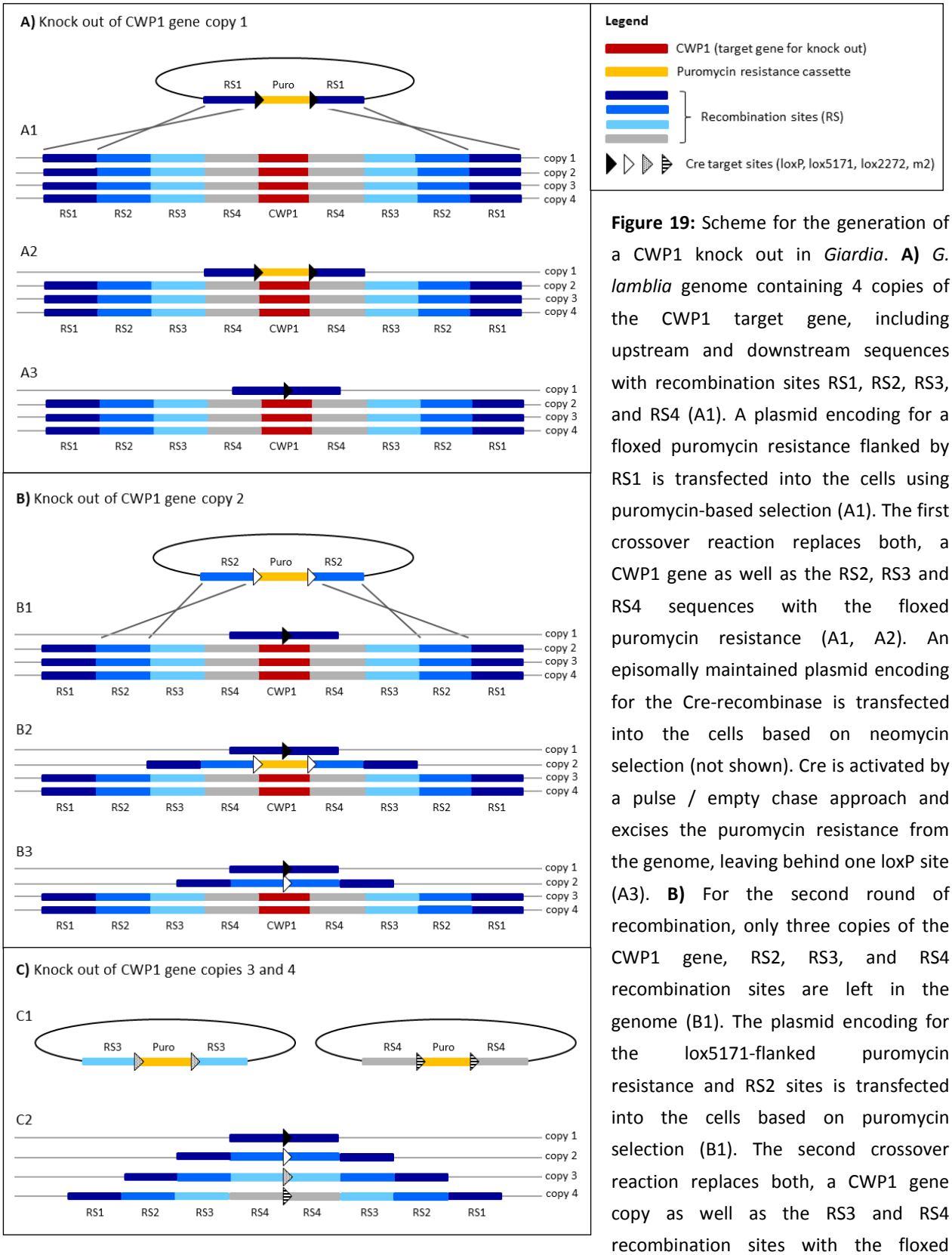


Figure 19: Scheme for the generation of a CWP1 knock out in *Giardia*. **A)** *G. lamblia* genome containing 4 copies of the CWP1 target gene, including upstream and downstream sequences with recombination sites RS1, RS2, RS3, and RS4 (A1). A plasmid encoding for a floxed puromycin resistance flanked by RS1 is transfected into the cells using puromycin-based selection (A1). The first crossover reaction replaces both, a CWP1 gene as well as the RS2, RS3 and RS4 sequences with the floxed puromycin resistance (A1, A2). An episomally maintained plasmid encoding for the Cre-recombinase is transfected into the cells based on neomycin selection (not shown). Cre is activated by a pulse / empty chase approach and excises the puromycin resistance from the genome, leaving behind one loxP site (A3). **B)** For the second round of recombination, only three copies of the CWP1 gene, RS2, RS3, and RS4 recombination sites are left in the genome (B1). The plasmid encoding for the lox5171-flanked puromycin resistance and RS2 sites is transfected into the cells based on puromycin selection (B1). The second crossover reaction replaces both, a CWP1 gene copy as well as the RS3 and RS4 recombination sites with the floxed puromycin resistance (B1, B2). Transfection of the Cre-encoding plasmid and activation of the enzyme leads to excision of the second CWP1 gene copy, leaving a lox5171 site behind (B3). For the third round of homologous

recombination, only two copies of CWP1 and two RS3 and RS4 recombination sites are left (B3). **C)** A third and fourth round of homologous recombination is performed by plasmids encoding for the puromycin resistance which is flanked by alternative loxP sites (lox2272 and m2) and RS3 and RS4 recombination sites, respectively (C1). The final product is a *G. lamblia* CWP1 knock out containing four alternative loxP sites where the CWP1 gene copies were removed (C2). Alternatively, the fourth round is performed without loxP-sites flanking the puromycin resistance, resulting in a *G. lamblia* CWP1 knock out with a stably integrated puromycin resistance cassette (not shown). However, to be able to perform functional analysis of CWP1 in a knock out strain, epitope-tagged wild-type or modified copies of the gene will have to be reinserted by yet another round of transfection.

2.2. Alternative applications of the Cre/loxP system in *G. lamblia*

Once established in *G. lamblia*, the Cre/loxP system provides a versatile tool to address diverse questions. The tool could be used as a sensitive detection system for rare events, for example to investigate if *G. lamblia* takes up DNA from the environment or is subject to horizontal gene transfer from microorganisms which inhabit the same ecological niche. An alternative approach focuses on the sexual behavior of *Giardia* which is a major point of discussion. The parasite has long been considered strictly asexual, but the presence of meiotic genes [122, 123] and increasing evidence from population genetic studies [124, 125] suggest that sexual recombination in fact occurs. However, this has not been shown *in vitro* so far.

To address these issues, we suggest designing two constructs which, if present in the same cell, lead to the establishment of a puromycin resistance (figure 20, A-C). Hereinafter we describe our approaches in more detail.

All suggested approaches are based on two constructs:

1. A first plasmid contains the ORF of the puromycin resistance cassette. In order to avoid expression of the *pac* gene at this stage, the start codon including promoter sequence is separated from the coding sequence by an intervening floxed neomycin resistance (figure 20A). Using neomycin-based selection the construct is stably integrated into the *G. lamblia* genome, giving rise to the *Giardia* strain termed “Start-Neofl-Puro”.
2. A second plasmid termed “Neo-pCWP1-Cre” encodes the Cre-recombinase with an upstream positioned inducible CWP1-promoter, and a neomycin resistance cassette (figure 20B).

As long as the two constructs are maintained in separate cells, the *pac* gene conferring puromycin resistance is not transcribed due to separation of the start codon and promoter sequence. However, as soon as the constructs are present in the same cell, activation of the Cre-recombinase by a pulse/empty chase approach (Wampfler and Hehl, manuscript in preparation) results in Cre-mediated excision of the floxed neomycin cassette (figure 20C). This places the puromycin start codon and the promoter sequence upstream of the puromycin ORF, resulting in puromycin-resistant parasites.

The establishment of an antibiotic resistance in the *Giardia* "Start-Neofl-Puro" strain upon acquisition of the "Neo-pCWP1-Cre" coding sequence allows in all approaches the positive selection of *Giardia* cells that **a)** took up the "Neo-pCWP1-Cre" DNA from the environment or the culture medium, respectively (figure 20 D-J), **b)** obtained DNA from co-cultured "Neo-pCWP1-Cre" plasmid expressing *E.coli* cells by horizontal gene transfer (not shown), and **c)** mediated sexual recombination with co-cultured *Giardia* cells that stably integrated the "Neo-pCWP1-Cre" sequence (figure 20 K-P). If occurring at all *in vitro*, these events are likely exceedingly rare, but because of the resulting resistance there is a good chance that they will be picked up by this sensitive system.

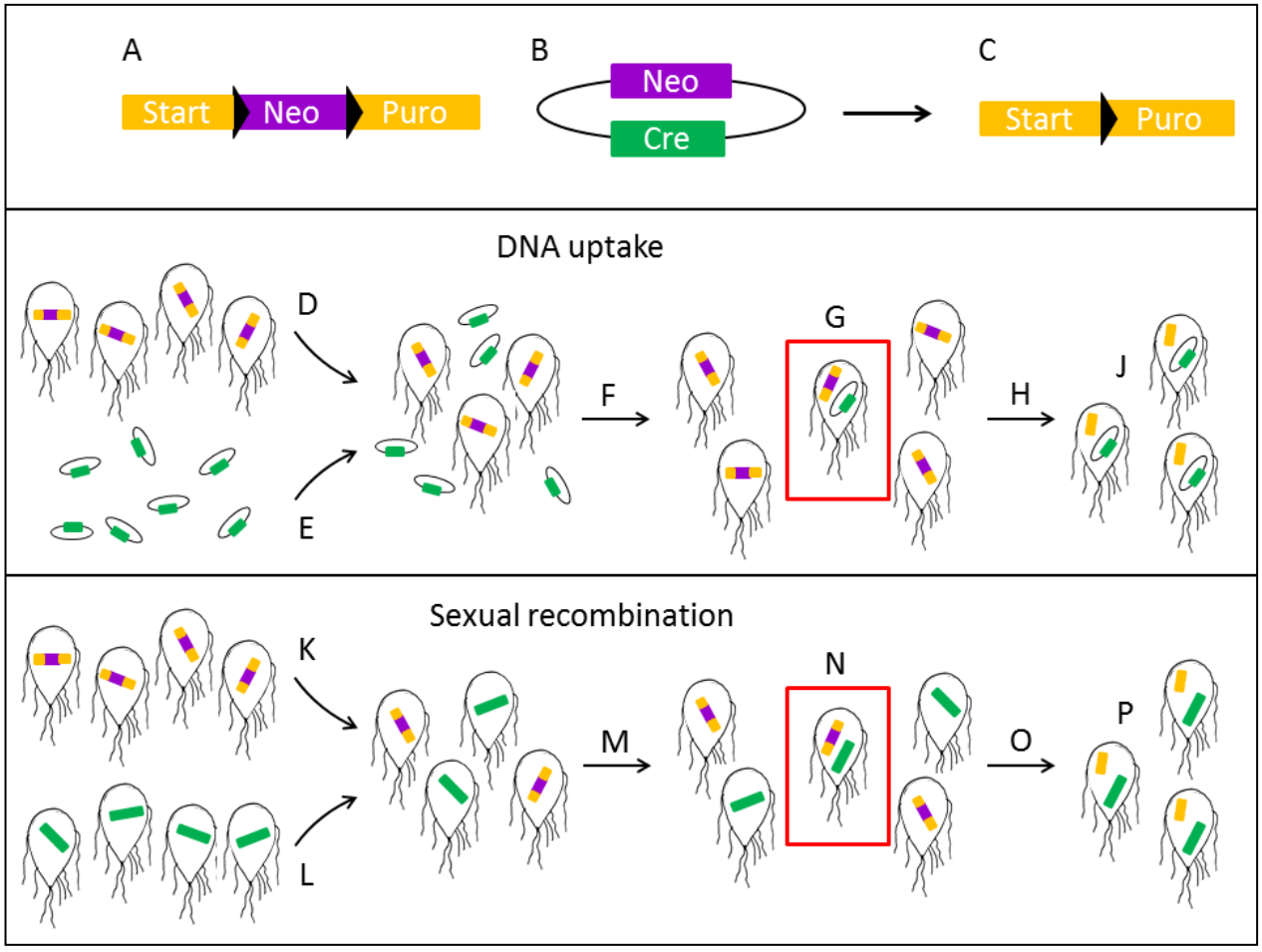


Figure 20: Cre/loxP system based approaches in *G. lamblia*. **A-C)** Constructs designed for the approaches. The first construct "Start-Neofl-Puro" (A) contains the ORF for a puromycin resistance cassette separated from its start codon and promoter sequence. The interrupting sequence is a floxed neomycin resistance cassette. The second construct "Neo-pCWP1-Cre" (B) contains the Cre-recombinase with an upstream pCWP1 promoter for inducible expression and a neomycin resistance cassette. If the two constructs (A and B) are maintained in the same cell, activation of the Cre-recombinase leads to excision of the floxed neomycin cassette and at the same time positioning of the puromycin start codon and promoter sequence right uptream of the puromycin ORF.
D-J) DNA uptake experiment. *G. lamblia* cells with stably integrated "Start-Neofl-Puro" (D) are grown with the "Neo-pCWP1-Cre" plasmid (E) in the culture medium. After some time of cultivation (F) some of the *Giardia* cells

might have taken up the DNA from the medium (G). Activation of the Cre-recombinase (H) results in the development of a puromycin resistance in these cells (J) which are selected by the addition of puromycin.
K-P) Sexual recombination experiment. *Giardia* cells with stably integrated "Start-Neofl-Puro" (K) and *Giardia* cells with stably integrated "Neo-pCWP1-Cre" (L) are cultivated together for a distinct time (M). If sexual recombination in vitro occurs, DNA is exchanged between few individual cells, resulting in few cells containing both constructs (N). Activation of the Cre-recombinase (O) results in the development of a puromycin resistance in these cells (P) which can be selected by the addition of puromycin.

3. References

1. Soltys, B.J., Falah, M. & Gupta, R.S. Identification of endoplasmic reticulum in the primitive eukaryote *Giardia lamblia* using cryoelectron microscopy and antibody to Bip. *J Cell Sci* **109 (Pt 7)**, 1909-17 (1996).
2. Stefanic, S., Palm, D., Svard, S.G. & Hehl, A.B. Organelle proteomics reveals cargo maturation mechanisms associated with Golgi-like encystation vesicles in the early-diverged protozoan *Giardia lamblia*. *J Biol Chem* **281**, 7595-604 (2006).
3. Jedelsky, P.L. et al. The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia* intestinalis. *PLoS One* **6**, e17285 (2011).
4. Morf, L. et al. The transcriptional response to encystation stimuli in *Giardia lamblia* is restricted to a small set of genes. *Eukaryot Cell* **9**, 1566-76 (2010).
5. Cao, Z. et al. Use of fluorescence-activated vesicle sorting for isolation of Naked2-associated, basolaterally targeted exocytic vesicles for proteomics analysis. *Mol Cell Proteomics* **7**, 1651-67 (2008).
6. Gauthier, D.J., Sobota, J.A., Ferraro, F., Mains, R.E. & Lazure, C. Flow cytometry-assisted purification and proteomic analysis of the corticotropes dense-core secretory granules. *Proteomics* **8**, 3848-61 (2008).
7. Brunner, Y., Schvartz, D., Coute, Y. & Sanchez, J.C. Proteomics of regulated secretory organelles. *Mass Spectrom Rev* **28**, 844-67 (2009).
8. Lee, Y.H., Tan, H.T. & Chung, M.C. Subcellular fractionation methods and strategies for proteomics. *Proteomics* **10**, 3935-56 (2010).
9. Jerlstrom-Hultqvist, J. et al. Hydrogenosomes in the diplomonad *Spironucleus* salmonicida. *Nat Commun* **4**, 2493 (2013).
10. Faso, C. & Hehl, A.B. Membrane trafficking and organelle biogenesis in *Giardia lamblia*: use it or lose it. *Int J Parasitol* **41**, 471-80 (2011).
11. Stefanic, S. et al. Neogenesis and maturation of transient Golgi-like cisternae in a simple eukaryote. *J Cell Sci* **122**, 2846-56 (2009).
12. Faso, C., Konrad, C., Schraner, E.M. & Hehl, A.B. Export of cyst wall material and Golgi organelle neogenesis in *Giardia lamblia* depend on endoplasmic reticulum exit sites. *Cell Microbiol* (2012).
13. Lanfredi-Rangel, A., Attias, M., Reiner, D.S., Gillin, F.D. & De Souza, W. Fine structure of the biogenesis of *Giardia lamblia* encystation secretory vesicles. *J Struct Biol* **143**, 153-63 (2003).
14. Marti, M. et al. An ancestral secretory apparatus in the protozoan parasite *Giardia* intestinalis. *J Biol Chem* **278**, 24837-48 (2003).
15. Hehl, A.B. & Marti, M. Secretory protein trafficking in *Giardia* intestinalis. *Mol Microbiol* **53**, 19-28 (2004).
16. Macechko, P.T., Steimle, P.A., Lindmark, D.G., Erlandsen, S.L. & Jarroll, E.L. Galactosamine-synthesizing enzymes are induced when *Giardia* encyst. *Mol Biochem Parasitol* **56**, 301-9 (1992).
17. Sener, K., Shen, Z., Newburg, D.S. & Jarroll, E.L. Amino sugar phosphate levels in *Giardia* change during cyst wall formation. *Microbiology* **150**, 1225-30 (2004).
18. Birkeland, S.R. et al. Transcriptome analyses of the *Giardia lamblia* life cycle. *Mol Biochem Parasitol* **174**, 62-5 (2010).
19. Faso, C., Bischof, S, and Hehl AB. The proteome landscape of *Giardia lamblia* encystation. *PLoS One* **in press** (2013).
20. Chiu, P.W., Huang, Y.C., Pan, Y.J., Wang, C.H. & Sun, C.H. A novel family of cyst proteins with epidermal growth factor repeats in *Giardia lamblia*. *PLoS Negl Trop Dis* **4**, e677 (2010).

21. Franzen, O. et al. Transcriptome profiling of *Giardia* intestinalis using strand-specific RNA-seq. *PLoS Comput Biol* **9**, e1003000 (2013).
22. Elde, N.C., Long, M. & Turkewitz, A.P. A role for convergent evolution in the secretory life of cells. *Trends Cell Biol* **17**, 157-64 (2007).
23. Dikeakos, J.D. & Reudelhuber, T.L. Sending proteins to dense core secretory granules: still a lot to sort out. *J Cell Biol* **177**, 191-6 (2007).
24. Orci, L. et al. Conversion of proinsulin to insulin occurs coordinately with acidification of maturing secretory vesicles. *J Cell Biol* **103**, 2273-81 (1986).
25. Chung, K.N., Walter, P., Aponte, G.W. & Moore, H.P. Molecular sorting in the secretory pathway. *Science* **243**, 192-7 (1989).
26. Cool, D.R. et al. Carboxypeptidase E is a regulated secretory pathway sorting receptor: genetic obliteration leads to endocrine disorders in Cpe(fat) mice. *Cell* **88**, 73-83 (1997).
27. Arvan, P. & Castle, D. Sorting and storage during secretory granule biogenesis: looking backward and looking forward. *Biochem J* **332 (Pt 3)**, 593-610 (1998).
28. Hehl, A.B., Marti, M. & Kohler, P. Stage-specific expression and targeting of cyst wall protein-green fluorescent protein chimeras in *Giardia*. *Mol Biol Cell* **11**, 1789-800 (2000).
29. Mouchantaf, R., Kumar, U., Sulea, T. & Patel, Y.C. A conserved alpha-helix at the amino terminus of prosomatostatin serves as a sorting signal for the regulated secretory pathway. *J Biol Chem* **276**, 26308-16 (2001).
30. Taupenot, L. et al. Identification of a novel sorting determinant for the regulated pathway in the secretory protein chromogranin A. *J Cell Sci* **115**, 4827-41 (2002).
31. Garcia, A.L. et al. A prohormone convertase cleavage site within a predicted alpha-helix mediates sorting of the neuronal and endocrine polypeptide VGF into the regulated secretory pathway. *J Biol Chem* **280**, 41595-608 (2005).
32. Lujan, H.D., Mowatt, M.R., Conrad, J.T., Bowers, B. & Nash, T.E. Identification of a novel *Giardia lamblia* cyst wall protein with leucine-rich repeats. Implications for secretory granule formation and protein assembly into the cyst wall. *J Biol Chem* **270**, 29307-13 (1995).
33. Reiner, D.S., McCaffery, J.M. & Gillin, F.D. Reversible interruption of *Giardia lamblia* cyst wall protein transport in a novel regulated secretory pathway. *Cell Microbiol* **3**, 459-72 (2001).
34. Davids, B.J., Mehta, K., Fesus, L., McCaffery, J.M. & Gillin, F.D. Dependence of *Giardia lamblia* encystation on novel transglutaminase activity. *Mol Biochem Parasitol* **136**, 173-80 (2004).
35. Konrad, C., Spycher, C. & Hehl, A.B. Selective condensation drives partitioning and sequential secretion of cyst wall proteins in differentiating *Giardia lamblia*. *PLoS Pathog* **6**, e1000835 (2010).
36. Reiner, D.S., McCaffery, M. & Gillin, F.D. Sorting of cyst wall proteins to a regulated secretory pathway during differentiation of the primitive eukaryote, *Giardia lamblia*. *Eur J Cell Biol* **53**, 142-53 (1990).
37. Beuret, N., Stettler, H., Renold, A., Rutishauser, J. & Spiess, M. Expression of regulated secretory proteins is sufficient to generate granule-like structures in constitutively secreting cells. *J Biol Chem* **279**, 20242-9 (2004).
38. Kim, T., Gondre-Lewis, M.C., Arnaoutova, I. & Loh, Y.P. Dense-core secretory granule biogenesis. *Physiology (Bethesda)* **21**, 124-33 (2006).
39. Voorberg, J. et al. Biogenesis of von Willebrand factor-containing organelles in heterologous transfected CV-1 cells. *EMBO J* **12**, 749-58 (1993).
40. Jarroll, E.L., Manning, P., Lindmark, D.G., Coggins, J.R. & Erlandsen, S.L. *Giardia* cyst wall-specific carbohydrate: evidence for the presence of galactosamine. *Mol Biochem Parasitol* **32**, 121-31 (1989).

41. Gerwig, G.J. et al. The *Giardia* intestinalis filamentous cyst wall contains a novel beta(1-3)-N-acetyl-D-galactosamine polymer: a structural and conformational study. *Glycobiology* **12**, 499-505 (2002).
42. Karr, C.D. & Jarroll, E.L. Cyst wall synthase: N-acetylgalactosaminyltransferase activity is induced to form the novel N-acetylgalactosamine polysaccharide in the *Giardia* cyst wall. *Microbiology* **150**, 1237-43 (2004).
43. Jarroll, E.L. et al. Regulation of carbohydrate metabolism during *Giardia* encystment. *J Eukaryot Microbiol* **48**, 22-6 (2001).
44. Lopez, A.B., Sener, K., Trosien, J., Jarroll, E.L. & van Keulen, H. UDP-N-acetylglucosamine 4'-epimerase from the intestinal protozoan *Giardia* intestinalis lacks UDP-glucose 4'-epimerase activity. *J Eukaryot Microbiol* **54**, 154-60 (2007).
45. Banerjee, S., Cui, J., Robbins, P.W. & Samuelson, J. Use of *Giardia*, which appears to have a single nucleotide-sugar transporter for UDP-GlcNAc, to identify the UDP-Glc transporter of *Entamoeba*. *Mol Biochem Parasitol* **159**, 44-53 (2008).
46. Chatterjee, A. et al. *Giardia* cyst wall protein 1 is a lectin that binds to curled fibrils of the GalNAc homopolymer. *PLoS Pathog* **6**, e1001059 (2010).
47. Midlej, V., Meinig, I., de Souza, W. & Benchimol, M. A New Set of Carbohydrate-positive Vesicles in Encysting *Giardia lamblia*. *Protist* (2012).
48. Hammarstrom, S., Murphy, L.A., Goldstein, I.J. & Etzler, M.E. Carbohydrate binding specificity of four N-acetyl-D-galactosamine- "specific" lectins: Helix pomatia A hemagglutinin, soy bean agglutinin, lima bean lectin, and Dolichos biflorus lectin. *Biochemistry* **16**, 2750-5 (1977).
49. McArthur, A.G. et al. The *Giardia* genome project database. *FEMS Microbiol Lett* **189**, 271-3 (2000).
50. Samuelson, J. et al. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proc Natl Acad Sci U S A* **102**, 1548-53 (2005).
51. Banerjee, S., Robbins, P.W. & Samuelson, J. Molecular characterization of nucleocytosolic O-GlcNAc transferases of *Giardia lamblia* and Cryptosporidium parvum. *Glycobiology* **19**, 331-6 (2009).
52. Morrison, H.G. et al. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* **317**, 1921-6 (2007).
53. Turkewitz, A.P. Out with a bang! *Tetrahymena* as a model system to study secretory granule biogenesis. *Traffic* **5**, 63-8 (2004).
54. Vayssie, L., Skouri, F., Sperling, L. & Cohen, J. Molecular genetics of regulated secretion in *Paramecium*. *Biochimie* **82**, 269-88 (2000).
55. Rosati, G. & Modeo, L. Extrusomes in ciliates: diversification, distribution, and phylogenetic implications. *J Eukaryot Microbiol* **50**, 383-402 (2003).
56. Hohl, H.R. & Hamamoto, S.T. Ultrastructure of spore differentiation in *Dictyostelium*: the prespore vacuole. *J Ultrastruct Res* **26**, 442-53 (1969).
57. Srinivasan, S., Alexander, H. & Alexander, S. Crossing the finish line of development: regulated secretion of *Dictyostelium* proteins. *Trends Cell Biol* **10**, 215-9 (2000).
58. Alexander, S., Srinivasan, S. & Alexander, H. Proteomics opens doors to the mechanisms of developmentally regulated secretion. *Mol Cell Proteomics* **2**, 1156-63 (2003).
59. Frisardi, M. et al. The most abundant glycoprotein of amebic cyst walls (Jacob) is a lectin with five Cys-rich, chitin-binding domains. *Infect Immun* **68**, 4217-24 (2000).
60. Van Dellen, K.L. et al. Unique posttranslational modifications of chitin-binding lectins of *Entamoeba invadens* cyst walls. *Eukaryot Cell* **5**, 836-48 (2006).
61. Chatterjee, A. et al. Evidence for a "wattle and daub" model of the cyst wall of *Entamoeba*. *PLoS Pathog* **5**, e1000498 (2009).

62. Mai, K. et al. Oocyst wall formation and composition in coccidian parasites. *Mem Inst Oswaldo Cruz* **104**, 281-9 (2009).
63. Dubremetz, J.F., Garcia-Reguet, N., Conseil, V. & Fourmaux, M.N. Apical organelles and host-cell invasion by Apicomplexa. *Int J Parasitol* **28**, 1007-13 (1998).
64. Ngo, H.M., Yang, M. & Joiner, K.A. Are rhoptries in Apicomplexan parasites secretory granules or secretory lysosomal granules? *Mol Microbiol* **52**, 1531-41 (2004).
65. Mercier, C., Adjogble, K.D., Daubener, W. & Delauw, M.F. Dense granules: are they key organelles to help understand the parasitophorous vacuole of all apicomplexa parasites? *Int J Parasitol* **35**, 829-49 (2005).
66. Carruthers, V.B.a.T., Fiona M. . Receptor-ligand interaction and invasion: Microneme proteins in apicomplexans. *Subcell Biochem* **47**, 33–45 (2008).
67. Scholtyseck, E., Mehlhorn, H. & Haberkorn, A. Die Feinstruktur der Makrogameten des Mäusecoccids *Eimeria* falciformis. *Zeitschrift für Parasitenkunde* **37**, 44-54 (1971).
68. Pittilo, R.M., Ball, S.J. & Hutchison, W.M. The ultrastructural development of the macrogamete of*Eimeria* stiedai. *Protoplasma* **104**, 33-41 (1980).
69. Ferguson, D.J., Belli, S.I., Smith, N.C. & Wallach, M.G. The development of the macrogamete and oocyst wall in *Eimeria* maxima: immuno-light and electron microscopy. *Int J Parasitol* **33**, 1329-40 (2003).
70. Tokuyasu, K. & Scherbaum, O.H. Ultrastructure of mucocysts and pellicle of *Tetrahymena* pyriformis. *J Cell Biol* **27**, 67-81 (1965).
71. Chilcoat, N.D., Melia, S.M., Haddad, A. & Turkewitz, A.P. Granule lattice protein 1 (Grl1p), an acidic, calcium-binding protein in *Tetrahymena* thermophila dense-core secretory granules, influences granule size, shape, content organization, and release but not protein sorting or condensation. *J Cell Biol* **135**, 1775-87 (1996).
72. Vazquez-Martinez, R. et al. Revisiting the regulated secretory pathway: from frogs to human. *Gen Comp Endocrinol* **175**, 1-9 (2012).
73. Chilcoat, N.D., Elde, N.C. & Turkewitz, A.P. An antisense approach to phenotype-based gene cloning in *Tetrahymena*. *Proc Natl Acad Sci U S A* **98**, 8709-13 (2001).
74. Cowan, A.T., Bowman, G.R., Edwards, K.F., Emerson, J.J. & Turkewitz, A.P. Genetic, genomic, and functional analysis of the granule lattice proteins in *Tetrahymena* secretory granules. *Mol Biol Cell* **16**, 4046-60 (2005).
75. Bowman, G.R., Smith, D.G., Michael Siu, K.W., Pearlman, R.E. & Turkewitz, A.P. Genomic and proteomic evidence for a second family of dense core granule cargo proteins in *Tetrahymena* thermophila. *J Eukaryot Microbiol* **52**, 291-7 (2005).
76. Verbsky, J.W. & Turkewitz, A.P. Proteolytic processing and Ca2+-binding activity of dense-core vesicle polypeptides in *Tetrahymena*. *Mol Biol Cell* **9**, 497-511 (1998).
77. Bradshaw, N.R., Chilcoat, N.D., Verbsky, J.W. & Turkewitz, A.P. Proprotein processing within secretory dense core granules of *Tetrahymena* thermophila. *J Biol Chem* **278**, 4087-95 (2003).
78. Hutton, J.C. *Tetrahymena*: the key to the genetic analysis of the regulated pathway of polypeptide secretion? *Proc Natl Acad Sci U S A* **94**, 10490-2 (1997).
79. Rahaman, A., Miao, W. & Turkewitz, A.P. Independent transport and sorting of functionally distinct protein families in *Tetrahymena* thermophila dense core secretory granules. *Eukaryot Cell* **8**, 1575-83 (2009).
80. Avron, B., Stolarsky, T., Chayen, A. & Mirelman, D. Encystation of *Entamoeba invadens* IP-1 is induced by lowering the osmotic pressure and depletion of nutrients from the medium. *J Protozool* **33**, 522-5 (1986).
81. Arroyo-Begovich, A., Carabez-Trejo, A. & Ruiz-Herrera, J. Identification of the structural component in the cyst wall of *Entamoeba invadens*. *J Parasitol* **66**, 735-41 (1980).

82. Ebert, F. et al. An *Entamoeba* cysteine peptidase specifically expressed during encystation. *Parasitol Int* **57**, 521-4 (2008).
83. Kim, J., Bae, S.S., Sung, M.H., Lee, K.H. & Park, S.J. Comparative proteomic analysis of trophozoites versus cysts of *Giardia lamblia*. *Parasitol Res* **104**, 475-9 (2009).
84. Gillon, A.D., Latham, C.F. & Miller, E.A. Vesicle-mediated ER export of proteins and lipids. *Biochim Biophys Acta* **1821**, 1040-9 (2012).
85. Siddiqi, S.A. VLDL exits from the endoplasmic reticulum in a specialized vesicle, the VLDL transport vesicle, in rat primary hepatocytes. *Biochem J* **413**, 333-42 (2008).
86. Siddiqi, S. et al. A novel multiprotein complex is required to generate the prechylomicron transport vesicle from intestinal ER. *J Lipid Res* **51**, 1918-28 (2010).
87. Saito, K. et al. TANGO1 facilitates cargo loading at endoplasmic reticulum exit sites. *Cell* **136**, 891-902 (2009).
88. Thor, F., Gautschi, M., Geiger, R. & Helenius, A. Bulk flow revisited: transport of a soluble protein in the secretory pathway. *Traffic* **10**, 1819-30 (2009).
89. Davids, B.J. et al. A new family of *Giardia*l cysteine-rich non-VSP protein genes and a novel cyst protein. *PLoS One* **1**, e44 (2006).
90. Touz, M.C. et al. The activity of a developmentally regulated cysteine proteinase is required for cyst wall formation in the primitive eukaryote *Giardia lamblia*. *J Biol Chem* **277**, 8474-81 (2002).
91. Touz, M.C., Lujan, H.D., Hayes, S.F. & Nash, T.E. Sorting of encystation-specific cysteine protease to lysosome-like peripheral vacuoles in *Giardia lamblia* requires a conserved tyrosine-based motif. *J Biol Chem* **278**, 6420-6 (2003).
92. DuBois, K.N. et al. Identification of the major cysteine protease of *Giardia* and its role in encystation. *J Biol Chem* **283**, 18024-31 (2008).
93. Davids, B.J. et al. An atypical proprotein convertase in *Giardia lamblia* differentiation. *Mol Biochem Parasitol* **175**, 169-80 (2011).
94. Kabnick, K.S. & Peattie, D.A. In situ analyses reveal that the two nuclei of *Giardia lamblia* are equivalent. *J Cell Sci* **95 (Pt 3)**, 353-60 (1990).
95. Bernander, R., Palm, J.E. & Svard, S.G. Genome ploidy in different stages of the *Giardia lamblia* life cycle. *Cell Microbiol* **3**, 55-62 (2001).
96. Yu, D.C., Wang, A.L. & Wang, C.C. Stable coexpression of a drug-resistance gene and a heterologous gene in an ancient parasitic protozoan *Giardia lamblia*. *Mol Biochem Parasitol* **83**, 81-91 (1996).
97. Sun, C.H., Chou, C.F. & Tai, J.H. Stable DNA transfection of the primitive protozoan pathogen *Giardia lamblia*. *Mol Biochem Parasitol* **92**, 123-32 (1998).
98. Singer, S.M., Yee, J. & Nash, T.E. Episomal and integrated maintenance of foreign DNA in *Giardia lamblia*. *Mol Biochem Parasitol* **92**, 59-69 (1998).
99. Davis-Hayman, S.R. & Nash, T.E. Genetic manipulation of *Giardia lamblia*. *Mol Biochem Parasitol* **122**, 1-7 (2002).
100. Touz, M.C., Conrad, J.T. & Nash, T.E. A novel palmitoyl acyl transferase controls surface protein palmitoylation and cytotoxicity in *Giardia lamblia*. *Mol Microbiol* **58**, 999-1011 (2005).
101. Touz, M.C., Gottig, N., Nash, T.E. & Lujan, H.D. Identification and characterization of a novel secretory granule calcium-binding protein from the early branching eukaryote *Giardia lamblia*. *J Biol Chem* **277**, 50557-63 (2002).
102. Lauwaet, T. et al. Protein phosphatase 2A plays a crucial role in *Giardia lamblia* differentiation. *Mol Biochem Parasitol* **152**, 80-9 (2007).
103. Prucca, C.G. et al. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* **456**, 750-4 (2008).

104. Touz, M.C., Kulakova, L. & Nash, T.E. Adaptor protein complex 1 mediates the transport of lysosomal proteins from a Golgi-like organelle to peripheral vacuoles in the primitive eukaryote *Giardia lamblia*. *Mol Biol Cell* **15**, 3053-60 (2004).
105. Rivero, M.R., Kulakova, L. & Touz, M.C. Long double-stranded RNA produces specific gene downregulation in *Giardia lamblia*. *J Parasitol* **96**, 815-9 (2010).
106. Carpenter, M.L. & Cande, W.Z. Using morpholinos for gene knockdown in *Giardia* intestinalis. *Eukaryot Cell* **8**, 916-9 (2009).
107. Castillo-Romero, A. et al. Rab11 and actin cytoskeleton participate in *Giardia lamblia* encystation, guiding the specific vesicles to the cyst wall. *PLoS Negl Trop Dis* **4**, e697 (2010).
108. Sternberg, N., Hamilton, D. & Hoess, R. Bacteriophage P1 site-specific recombination. II. Recombination between loxP and the bacterial chromosome. *J Mol Biol* **150**, 487-507 (1981).
109. Austin, S., Ziese, M. & Sternberg, N. A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell* **25**, 729-36 (1981).
110. Hoess, R.H., Ziese, M. & Sternberg, N. P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc Natl Acad Sci U S A* **79**, 3398-402 (1982).
111. Barrett, B., LaCount, D.J. & Donelson, J.E. *Trypanosoma brucei*: a first-generation CRE-loxP site-specific recombination system. *Exp Parasitol* **106**, 37-44 (2004).
112. Brecht, S., Erdhart, H., Soete, M. & Soldati, D. Genome engineering of *Toxoplasma gondii* using the site-specific recombinase Cre. *Gene* **234**, 239-47 (1999).
113. O'Neill, M.T., Phuong, T., Healer, J., Richard, D. & Cowman, A.F. Gene deletion from *Plasmodium falciparum* using FLP and Cre recombinases: implications for applied site-specific recombination. *Int J Parasitol* **41**, 117-23 (2011).
114. Dan, M., Wang, A.L. & Wang, C.C. Inhibition of pyruvate-ferredoxin oxidoreductase gene expression in *Giardia lamblia* by a virus-mediated hammerhead ribozyme. *Mol Microbiol* **36**, 447-56 (2000).
115. Branda, C.S. & Dymecki, S.M. Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* **6**, 7-28 (2004).
116. Andenmatten, N. et al. Conditional genome engineering in *Toxoplasma gondii* uncovers alternative invasion mechanisms. *Nat Methods* **10**, 125-7 (2013).
117. Banaszynski, L.A., Liu, C.W. & Wandless, T.J. Characterization of the FKBP.rapamycin.FRB ternary complex. *J Am Chem Soc* **127**, 4715-21 (2005).
118. Hoess, R.H., Wierzbicki, A. & Abremski, K. The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res* **14**, 2287-300 (1986).
119. Lee, G. & Saito, I. Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene* **216**, 55-65 (1998).
120. Langer, S.J., Ghafoori, A.P., Byrd, M. & Leinwand, L. A genetic screen identifies novel non-compatible loxP sites. *Nucleic Acids Res* **30**, 3067-77 (2002).
121. Wall, R.J. New gene transfer methods. *Theriogenology* **57**, 189-201 (2002).
122. Ramesh, M.A., Malik, S.B. & Logsdon, J.M., Jr. A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol* **15**, 185-91 (2005).
123. Birky, C.W., Jr. Sex: is *Giardia* doing it in the dark? *Curr Biol* **15**, R56-8 (2005).
124. Cooper, M.A., Adam, R.D., Worobey, M. & Sterling, C.R. Population genetics provides evidence for recombination in *Giardia*. *Curr Biol* **17**, 1984-8 (2007).
125. Logsdon, J.M., Jr. Evolutionary genetics: sex happens in *Giardia*. *Curr Biol* **18**, R66-8 (2008).

Acknowledgements

I would like to thank all people who contributed to this work or supported me in any other way during the time of my PhD studies at the Institute of Parasitology in Zurich.

First, I would like to thank my supervisor Prof. Dr. Adrian Hehl for giving me the opportunity to work on interesting, challenging and manifold projects. I thank him for his constant and valuable support, for the many things he taught me, and for all the interesting biological discussions. I appreciated his open-minded attitude for novel ideas and projects, and his way of supervising that promoted my independence.

I thank Dr. Cornelia Spycher and Dr. Carmen Faso for great and pleasant teamwork, for critical and helpful discussions, as well as for valuable support and advice in all areas of my work.

A thousand thanks go to Therese Michel for cloning of many constructs. I thank her for all the valuable input, for a critical eye in the right situations, and for her endurance and tenacity when she had to clone constructs of the highest complexity.

I would like to thank Dr. Sasa Stefanic, Vinko Tosevski, Dr. Christoph Lippuner and Dr. Claudia Dumrese for helping me with the flow cytometer and the confocal microscope day and night.

Many thanks go to my PhD thesis committee members Prof. Dr. François Verrey, Prof. Dr. Thorsten Hornemann and Dr. Werner Kovacs for supervising my thesis.

Special thanks go to the head of the institute, Prof. Dr. Peter Deplazes, as well as to the administration and secretarial team for their support in many organisatorial things.

I would like to thank all members of the Hehl group and of the Institute of Parasitology who made the time of my PhD unforgettable. I highly appreciated the entertaining chats with Dr. Jon Paulin Zumthor and his conspicuous music collection, especially when combined with Samuel Rout's snappy dance interludes. I thank our team for making lab life funny even in rather frustrating lab days – they sometimes made me laugh till I cried (I will never forget the "whatever box" in our fridge). I thank Dr. Carmen Faso and Jacqueline Ebneter for taking over my favorite projects and I wish them good luck with further research.

Last but not least I thank my family and my friends for always being there for me during my PhD years. I thank them for supporting me in cursing or celebrating, for going frustration-shopping or a glass of wine when required, for entertaining me during overnight shifts in the laboratory, for help in statistics, and for good advice in difficult situations.

Curriculum Vitae

PERSONAL DATA

Last name	WAMPFLER
First name	Petra Barbara
Date of birth	21st of March 1982
Citizenship	Diemtigen BE, Switzerland
Address	Murhaldenweg 16 8057 Zurich
Contact details	Mobile: +41 (0) 78 769 93 62 petra.wampfler@access.uzh.ch

CAREER

11/2009 – today **Institute of Parasitology, University of Zurich**

PhD student in the Molecular Parasitology group of Prof. Adrian B. Hehl

- Project 1: Organelle proteomics in *Giardia lamblia*
- Project 2: Implementation of the Cre/loxP system in *Giardia lamblia*
- Supervision of a Masters student in their lab work (1 year)
- Assistance in organisation and delivery of four course components

05/2008 – 11/2009 **Baxter AG, Volketswil, Canton Zurich**

Regulatory Affairs / Pharmacovigilance Assistant

10/2004 – 06/2006 **Institute of Biochemistry, University of Zurich**

Employee in the analytical laboratory for genotyping
Workload of 40% alongside my studies at ETH Zurich

11/2001 – 06/2002 **St. Gallische Creditanstalt, St. Gallen**

Customer Services

EDUCATION

10/2002 – 03/2008 **Diploma in Integrative Biology at ETH Zurich (MNF, D-BIOL)**

Issued: 13th of March 2008

Basic studies: Biology (with chemistry as a major component)

Professional studies: Integrative biology incorporating five examined units:
pharmacology, toxicology, biochemistry, cell biology,
food microbiology (diploma thesis)

Diploma Thesis: "Construction of *L. innocua* CLIP11262*, L. ivanovii* WSLC 3009 and *L. welshimeri* WSLC 5008 Deletion Mutants and Analysis of HGFP-CBD500 Binding Properties"
- Laboratory of Food Microbiology, ETH Zurich
- Prof. Dr. Martin J. Loessner

08/1997 – 08/2001 **High school diploma at Kantonsschule am Burggraben St. Gallen**

Mathematics and natural sciences (Type C)

PUBLICATIONS

Under review:	Proteomics of secretory and endocytic organelles in *Giardia lamblia* **Petra B. Wampfler**, Vinko Tosevski, Paolo Nanni, Cornelia Spycher, Adrian B. Hehl
In preparation:	The Cre/loxP system in *Giardia lamblia*: genomic manipulations in a tetraploid protozoan **Petra B. Wampfler**, Carmen Faso, Adrian B. Hehl

COURSES / FURTHER EDUCATION

Genomics/Proteomics	Lecture "Functional Genomics" in Zurich (FS/2012) Lecture "Proteomics in Biology" in Zurich (FS/2011) Workshop "Protein Identification by Mass Spectrometry" in Lausanne (2010)
Flow Cytometry	BD FACSAria I and III: independent user at the Flow Cytometry Facility Zurich BC Gallios, various training sessions
Other workshops	Open Access: Improvement of Scientific Publishing (09/2010) Scientific Writing in the Sciences or Medicine (10/2010) The Successful Start of a Professional Career (02/2011) Effective Scientific Presentation and Confident Public Speaking (03/2011) Projektmanagement für die Forschung (11/2011) Winterschool Microscopy Course (01/2012)

CONGRESSES ATTENDED

2012	**Annual Molecular Parasitology Meeting (MPM), Woods Hole MA, USA** Poster: "Proteomics of Secretory Organelles in *Giardia lamblia*"
2012	**Joint Annual Meeting (SSTMP, SSI, SSHH, SSM), St. Gallen, CH** Poster: "Proteomics of secretory organelles in *Giardia lamblia"*
2010-2012	**Swiss Society of Tropical Medicine and Parasitology (SSTMP) Student Meeting** Presentation: "Proteomics of secretory organelles in *G. lamblia"* (2012, 2011) Presentation: "Lost in evolution – a stage-regulated Golgi-remnant in the protozoan parasite *Giardia lamblia"* (2010)

AWARDS

09/2012	**2012 Elsevier Press Poster Award** Poster: "Proteomics of secretory organelles in *Giardia lamblia"* Presented at: Molecular Parasitology Meeting (MPM) 2012, Woods Hole, MA, USA

RESEARCH FUNDING

07/2011	"**Forschungskredit 2011" awarded by the University of Zurich (50 820 CHF)** Project: „Lost in evolution – a stage-regulated Golgi-remnant in the protozoan parasite *Giardia lamblia"*